# Labos

Revista de Derecho del Trabajo y Protección Social

Noviembre 2024

Volumen 5 - Número extraordinario


## Normativa europea sobre inteligencia artificial




uc3m | Universidad Carlos III de Madrid
Vicerrectorado de Política Científica

EISSN: 2660-7360
www.uc3m.es/labos

**Labos. Revista de Derecho del Trabajo y Protección Social**
EISSN: 2660-7360 - www.uc3m.es/labos



Labos. Revista de Derecho del Trabajo y Protección Social
Secretaría editorial
Universidad Carlos III de Madrid
c/ Madrid-126 28903 Getafe (Madrid) España
Correo electrónico: revistalabos@uc3m.es

# PRESENTACIÓN

## Economía de datos, algoritmos productivos y extractivos, y riesgos sociolaborales

## Data economy, productive and extractive algorithms, and socio-occupational risks

Ignasi Beltrán de Heredia Ruiz
*Catedrática de Derecho del Trabajo y de la Seguridad Social*
*Universitat Oberta de Catalunya (UOC)*

Adrián Todolí Signes
*Profesor Titular de Derecho del Trabajo y de la Seguridad Social*
*Universitat de València Estudi General*

doi: 10.20318/labos.2024.9028

Uno de los factores recurrentes para medir el progreso humano es la evolución de la técnica. Aunque se trata de un parámetro reduccionista, pues, no todas las grandes innovaciones se han materializado en forma de *herramientas* (de hecho, las más importantes –como las *ideas* o los *conceptos*–, no lo son), lo cierto es que, en las últimas décadas, el avance tecnológico ha sido prodigioso.

El detonador de esta aceleración ha sido, fundamentalmente, el efecto combinado de dos elementos: datos e incremento en varios órdenes de magnitud de la capacidad de procesamiento. La adquisición de *conocimiento*, que la digitalización y su computarización posibilitan, ha convertido a los datos en el nuevo polvo de oro para fundir lingotes. De ahí que un sector creciente e influyente de la economía haya pasado de producir bienes basados en átomos a bienes basados en bytes.

Esta nueva *economía de datos* ha impulsado una desenfrenada carrera hacia la datificación masiva, con importantes efectos epistemológicos. Si la digitalización permite tabular y procesar enormes cantidades de datos, ya no tiene sentido recurrir a una muestra (tal y como la estadística frecuentista exige). En este nuevo contexto basado fundamentalmente en correlaciones (y en el que a la causalidad queda relegada a un plano secundario o marginal), de lo que se trata es de recopilar todos los datos posibles y, cuando sea factible, absolutamente todos (incluso los erróneos).

Esta portentosa estadística computacional (todavía lejos de la *inteligencia* en el sentido humano del término) aspira a mejorar la detección de patrones y, en último término,

superar la capacidad humana de hacer pronósticos. En definitiva, como si el problema de la inducción hubiera dejado de existir y también se hubiera disipado el ruido estadístico, de lo que se trata es de *cartografiar* minuciosamente el pasado, para proyectar prístinamente el futuro a través de la ley de los grandes números. A su vez, la capacidad de identificar regularidades también ha permitido alcanzar cotas hace décadas inimaginables, pues, las máquinas ya son capaces de *mimetizar* facultades hasta hace poco reservadas al inconsciente humano (como la detección de imágenes y/o la generación de lenguaje complejo).

Las implicaciones de este proceso son abisales y empresarios y trabajadores difícilmente podrán resistirse al empuje de este avance tecnológico. Respecto de los primeros, en la medida que la identificación de patrones con fines predictivos puede otorgar una ventaja competitiva, las organizaciones empresariales se encuentran ante un punto de inflexión: su supervivencia pasará preeminentemente por la parametrización del mayor número de factores que intervienen en el proceso productivo y con la mayor granularidad que sea posible. En efecto, el uso de estos algoritmos extractivos hará que las empresas estén preeminentemente dirigidas *por los datos* (*data driven manufacture*). Especialmente porque, a medida que la detección de regularidades mejore, la analítica predictiva permitirá reducir el umbral de incertidumbre. La cuenca de atracción de este proceso será tan poderosa que (como se expone en la obra *Algoritmos productivos y extractivos. Cómo regular la digitalización para mejorar el empleo e incentivar la innovación*, Aranzadi, 2023) empujará a las empresas a delegar en las máquinas la evaluación de los riesgos y/o la toma de decisiones. Esta progresiva *desconexión de la inteligencia de la conciencia* será apreciable en toda la secuencia del programa de prestación contractual (incluyendo las fases previas) y, de hecho, ya se está empezando a manifestar a través de decisiones semiautomáticas (requiriendo la intervención humana en una segunda instancia) o directamente automáticas.

El impacto de esta gran delegación empresarial en las personas trabajadoras, por su parte, describirá una nueva y más aguda versión científica del trabajo, describiendo inquietantes y novedosas amenazas. La irrupción de asistentes informacionales en los entornos laborales es hoy por hoy creciente, especialmente porque, salvo en su versión más extrema de sustitución de la fuerza de trabajo, el *conocimiento* que se desprende de su capacidad de predicción estadística empieza a despuntar como el mejor complemento para superar las sistemáticas limitaciones cognitivas del ser humano (profusamente analizadas por la psicología de la conducta y la neurociencia). Este creciente *gobierno de los números* (en términos de Alain SUPIOT) acabará redefiniendo el rol de los trabajadores: por un lado, asumiendo un papel eminentemente reactivo a estímulos pronosticadores que, en forma de información y/o decisión, serán sido *cocinados* por algoritmos; y, al mismo tiempo, sometiéndolas a una pegajosa evaluación exhaustiva de los logros mediante indicadores numéricos. Cada miembro de la plantilla deberá responder al imperativo del individuo programado. De modo que, evidenciando un cambio radical en la concepción *hombre-máquina*, deberá objetivarse a sí mismo a través de la alineación de su rendimiento al patrón dictado por las máquinas (provocando, incluso, una permutación del binomio *máquina-hombre*).

La desencriptación de los procesos mentales más profundos (permitiendo el acceso a estratos de nuestro ser que, trascendiendo la percepción consciente, se encuentran

a diversas brazadas de profundidad) es el siguiente estadio a la vuelta de la esquina. La psicometría humana que la constante interacción con algoritmos extractivos posibilita, ya está permitiendo el acceso al *patio trasero neuronal* y, a través de correlaciones, a los procesos mentales más insondables que rigen el comportamiento y el pensamiento humano. La creciente proliferación de interfaces cerebro-ordenador en el entorno laboral (en este primer estadio, en aras a la preservación de la salud y la prevención de riesgos laborales) permitirá sumergirse por debajo del nivel consciente, superando la capacidad de autopercepción de uno mismo. El propósito en esta caza de los llamados *neurodatos* es simple (y, particularmente inquietante): se aspira al condicionamiento efectivo y fluido de la conducta. Este nuevo poder empresarial instrumentario, a través del acceso a los procesos inconscientes que rigen el comportamiento de los trabajadores, podrá alcanzar una condición de certeza *sin resistencia*, en forma de resultados garantizados. Un poder empresarial *de dirección* tan efectivo que el *de control* pasaría a ser marginal. En esta encrucijada (abordada en la obra *Inteligencia artificial y neuroderechos: la protección del yo inconsciente de la persona*, Aranzadi, 2023), todo apunta a que la preservación del yo insconsciente de las personas trabajadoras se erigirá en una de las metas ineludibles a alcanzar por el iuslaboralismo de cualquier democracia liberal.

A la luz de todo lo expuesto, las normas sociolaborales se encuentran frente a un escenario absolutamente disruptivo y el reto jurídico-positivo que plantea es mayúsculo y hercúleo.

El número especial de la revista que nos complace presentar es uno de los frutos del proyecto de investigación del Ministerio de Ciencia e Innovación, titulado "Algoritmos extractivos y neuroderechos. Retos regulatorios de la digitalización del trabajo". En concreto, recoge las ponencias del *Congreso sobre la IA en el mundo del trabajo* (celebrado en Valencia, los días 20 y 21 de junio 2024[1]). Este encuentro, que contó con la participación de destacados integrantes del mundo académico y profesional, tuvo por objeto el análisis del impacto del Reglamento de Inteligencia Artificial y de la nueva Directiva de plataformas digitales en el mundo laboral. Los directores del Congreso, firmantes de esta editorial, consideramos que la calidad de las ponencias, el interés de las mismas y la actualidad del tema –dada la reciente aprobación de la normativa analizada– merecían la publicación de las mismas. Agradecemos a los editores de la Revista, los prof. Mercader Uguina, Goerlich Peset y de la Puebla Pinilla la aceptación para realizar un número especial con las ponencias.

Las comunicaciones del Congreso pueden encontrarse en un libro publicado por el editorial Aranzadi y que en conjunto a este número especial de la revista Labos forma parte de los resultados del Congreso. En las aportaciones de ambos volúmenes (este libro y el número especial de la revista Labos) se desgranan las claves de muchos de estos desafíos y constituyen una aproximación de primer orden y de vanguardia de un futuro cuyas amenazas ya empiezan a materializarse.

[1] Las ponencias se pueden ver grabadas en el siguiente enlace https://adriantodoli.com/2024/07/08/videos-de-las-ponencias-del-congreso-de-ia-y-trabajo-celebrado-en-valencia-en-20-y-21-de-junio/

# ARTÍCULOS

# Obligaciones de transparencia y protección de datos en el ámbito de las relaciones laborales

## Transparency duties and data protection in the context of labour relations

Eva María Blázquez Agudo
*Catedrática Derecho Trabajo y Seguridad Social*
*Universidad Carlos III de Madrid*
ORCID ID: 0000-0002-8214-1960

doi: 10.20318/labos.2024.9029

*Resumen:* La relación del principio de transparencia y la protección de datos es bidireccional; se complementan y se limitan. En este trabajo, se analiza como el deber empresarial de información a las personas trabajadoras sobre el tratamiento de sus datos personales determina su licitud. Asimismo, se examinan los límites que la protección de datos impone a la información que recibe la representación de las personas trabajadoras, siempre sin impedir el ejercicio de sus funciones. A continuación, se estudia cómo el principio de transparencia se aplica a las decisiones automatizadas en la empresa, en especial su regulación en el RGPD. Por último, se aboga por el derecho de explicación como un grado superior al derecho a la información a los efectos de conseguir la transparencia.
*Palabras clave:* Protección de datos, principio de transparencia, derecho a la información, derecho de explicación

*Abstract:* The relationship between the transparency principle and data protection is bidirectional, both complement and limited each other. In this paper, how the company's duty to inform workers about the processing of their personal data to determine the lawfulness of such processing is analysed. Moreover, the limits that data protection imposes in the field of information received by the workers' representation is also studied. It is necessary to prevent that their functions are not impeded. After this, how the transparency principle is applied to automated decisions in the company is examined, especially its regulation in the GDPR. Finally, the right to explanation is advocated as a higher level than the right to information in the purposes of achieving transparency.
*Keywords:* Data protection, transparency principle, right to information, right to explanation

*Evamaria.blazquez@uc3m.es

## I. Breve introducción a la transparencia en la empresa

La transparencia ha sido calificada como un término *polisémico*, sobre el cual es difícil establecer una definición unívoca, la cual dependerá, en cierto modo, de la perspectiva que se elija en su análisis. Pero, además, es esencialmente *instrumental*, en cuanto a que la transparencia tiene valor como instrumento para detectar conductas inapropiadas y ejercer frente a ellas derechos de resarcimiento[1].

De acuerdo con la Real Academia Española a la Lengua se entiende por "*transparencia",* la cualidad de transparente, como sinónimo de *luminosidad*, *claridad, evidencia*, entre otros. Y, por "*transparente"*, *aquel cuerpo que permite ver los objetos con nitidez a través del él; aquello que es claro y evidente, que se comprende sin duda ni ambigüedad.* Para añadir posteriormente que se entiende como transparente a *aquellas instituciones o entidades que proporcionan información suficiente sobre su manera de actuar.*

De forma que las empresas serán transparentes cuando proporcionen información suficiente sobre su forma de actuación, información que servirá para el control legal de su actividad. De esta afirmación se deriva que el principio de transparencia se manifiesta principalmente a través de la exigencia de la transmisión de información sobre determinadas cuestiones a diferentes agentes con el fin de comprobar la adecuada actuación de la empresa.

En el ámbito de la empresa, el principio de transparencia se desarrolla fundamentalmente a través del derecho de información, donde cada vez se amplían más las exigencias legales sobre los datos que debe suministrar, tratando de verificar que se están respetando los derechos laborales de las personas trabajadoras.

En este contexto, por tanto, uno de los principales objetivos de la transparencia en el ámbito de las relaciones laborales será la búsqueda del equilibrio entre el poder de dirección y los derechos fundamentales de las personas trabajadoras. Desde esta perspectiva el principio de transparencia es un elemento instrumental que garantiza la aplicación del derecho al acceso a la justicia recogido en el artículo 24 de la Constitución en cuanto a que la información recibida va a servir para conocer si han sido vulnerados los derechos de las personas trabajadoras y, en su caso, en qué medida, buscando facilitar su capacidad de respuesta ante estas situaciones.

En todo caso, más allá de imponer obligaciones, la aplicación del principio de transparencia es una herramienta eficaz para la empresa en cuanto a que impulsa la confianza de sus clientes y proveedores, pero también de la propia Administración pública y de las personas trabajadoras[2]. Así como tiene la función de promover una mayor reputación corporativa que incide en la mejora de su negocio.

---

[1] COTINO HUESO, Lorenzo, "Transparencia y explicabilidad de la inteligencia artificial y "compañía" (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad…). Para qué, para quién y cuánta", EN: *Transparencia y explicabilidad de la IA*, Tirant, 2022.

[2] SÁNCHEZ DE DIEGO FERNÁNDEZ DE LA RIVA, Manuel, "Prólogo", EN: *Apuntes sobre transparencia*, Universidad Complutense de Madrid, 2018, pp. 9 y 10.

## II. La relación bidireccional entre protección de datos personales y el principio de transparencia

En general, pero también en el ámbito de la empresa, las normas sobre protección de datos personales acotan el principio de transparencia en cuanto a que limitan la obligación empresarial de suministrar información, tanto a las personas trabajadoras como a su representación; pero, por otra parte, el derecho a la información de las personas afectadas por el tratamiento de sus datos personales tiene un importante rol como elemento de seguridad en el desarrollo de dicha protección. En definitiva, ambos se limitan y complementan, siendo fundamental la búsqueda del punto de equilibrio entre ellos.

El principio de transparencia se califica como elemento fundamental del desarrollo de la protección de datos personales. Varios son los reconocimientos legales en este sentido. En primer lugar, el artículo 5 del Reglamento (UE) 2016/679, del Parlamento Europeo y del Consejo 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 y por el que se deroga la Directiva 95/46/CE (en adelante, RGPD), determina que a todos los tratamientos de datos personales se deben aplicar determinados principios como son los principios de licitud, lealtad y transparencia. Así, el principio de transparencia impone una obligación directa a través del derecho de información que se reconoce a las personas afectadas por dicho tratamiento.

Asimismo, en sentido similar, pero ya en el ámbito de las relaciones laborales, el artículo 88 del mismo cuerpo legal proclama que por ley o convenio colectivo se puede ampliar la protección de los derechos en relación a la protección de datos personales de las personas trabajadoras. En su párrafo segundo, se añade que se podrán establecer medidas adecuadas y específicas para preservar la dignidad humana de las personas interesadas, así como sus intereses legítimos y sus derechos fundamentales, prestando especial atención a la transparencia del tratamiento.

Ya en el ámbito nacional, la Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales (en adelante, LOPD), en su Título II, el capítulo I, se titula "*Transparencia e información*"; y, en concreto, su artículo 11 se denomina "*Transparencia e información al afectado*". En definitiva, se reconoce expresamente que la obligación de información sobre los datos personales está ligada a la idea de transparencia.

## III. El derecho individual de información como límite del tratamiento de datos personal

Dentro de esta relación entre ambos elementos (protección de datos personales y transparencia), es fundamental el análisis de como el derecho de información de la persona afectada por el tratamiento de los datos personales es una manifestación básica del principio de transparencia, en especial en el ámbito de las relaciones laborales.

En el tratamiento de datos personales, es esencial el consentimiento de la persona afectada para que sea licito, no obstante, el artículo 6.1 del RGPD recoge ciertas excepciones. Se reconoce que la empresa puede tratar los datos de las personas trabajadoras, sin su consentimiento, siempre que dicho tratamiento se realice en el marco del contrato laboral y solo para el cumplimiento de las obligaciones derivadas de la relación de trabajo.

En conclusión, es el propio contrato laboral la base legítima para el tratamiento de los datos personales de las personas trabajadoras en el normal desenvolvimiento de las relaciones en la empresa.

En cualquier caso, con independencia de esta excepción, la persona trabajadora deberá ser informada del tratamiento de sus datos personales, sustituyendo dicha información al consentimiento, y cumpliendo el principio de transparencia, sobre todo en los casos en que exista peligro de vulneración de sus derechos fundamentales.

Con el objeto de respetar este derecho de información de las personas trabajadoras se debe seguir los preceptos que se regulan sobre dicha cuestión en el RGPD y, en su caso, en la LOPD. Aunque hay que advertir que no existe una regulación concreta aplicable a las relaciones laborales, sino que habrá que adaptar la normativa general a las necesidades del derecho del trabajo.

En este análisis, hay que partir del Considerando 60 del RGPD, donde se señala que el responsable del tratamiento de los datos personales deberá aportar toda la información complementaria que sea necesaria para garantizar un tratamiento leal y transparente, con el fin de cumplir el principio de transparencia[3].

## 1. Las características de la información

El Considerando 58 del RGPD declara que a través del principio de transparencia se exige que toda información sea concisa, fácilmente accesible y fácil de entender, y que se utilice un lenguaje claro y sencillo, en especial en situaciones en las que la proliferación de agentes y la complejidad tecnológica de la práctica hagan que sea difícil para la persona interesada saber y comprender si se están recogiendo, por quién y con qué finalidad, datos personales que le conciernen.

En definitiva, el principio de transparencia exige a las personas responsables y a las encargadas del tratamiento de los datos personales que la información suministrada, a las personas cuyos datos personales sean tratados, sea clara y con un contenido entendible[4], de acuerdo con el artículo 12 del RGPD que describe la información como concisa, transparente, inteligible y de fácil acceso, con un lenguaje claro y sencillo.

[3] AEPD/ APDCAT/ Agencia Vasca de Protección de Datos, *Guía para el cumplimiento del deber de informar*, disponible en: https://www.aepd.es/documento/guia-modelo-clausula-informativa.pdf (consulta:28/08/2024), p. 2.

[4] Grupo de trabajo sobre protección de datos del Art. 29 (GT29), *Directrices sobre la transparencia en virtud del Reglamento (UE) 2016/679,* adoptadas el 29 de noviembre de 2017, revisadas por última vez y adoptadas el 11 de abril de 2018, p. 6.

## 2. Los medios de transmisión de la información

Las personas responsables del tratamiento de los datos personales, en este caso, las empresas, deberán tener especial cuidado en la elección de los medios de comunicación empleados para la trasmisión de la información a las personas interesadas[5].

En este sentido, el artículo 12 del RGPD señala que la información será facilitada por escrito o, si procede, por medios electrónicos. Así, la empresa podrá poner la información a disposición de las personas trabajadoras a través de formularios de papel o colgados en la web, registros en teléfonos móviles o datos de sensores. Si se trata de datos personales que no ha aportado la persona interesada, la comunicación sobre la información del tratamiento podrá enviarse por correo postal, mail o notificación vía intranet[6].

De hecho, el propio RGPD señala que en el caso de que la información sea solicitada por las personas interesadas por medios electrónicos, la respuesta, en la medida que sea posible, deberá ser también a través de estos medios. Pero, asimismo, cuando así lo solicite la persona trabajadora, la información podrá facilitarse verbalmente siempre que se demuestre la identidad de dicha persona.

## 3. El plazo de la información

En cuanto al plazo, el Considerando 61 del RGPD señala que se debe facilitarse a las personas interesadas la información sobre el tratamiento de sus datos personales en el momento en que se obtengan directamente de ellas o, si se consiguen a través de otra fuente, en un plazo razonable, dependiendo de las circunstancias del caso.

Si los datos personales son aportados por las personas interesadas, en el momento en que se obtengan, habrá que informarles sobre el tratamiento, siempre que no dispongan ya de dicha información.

En el caso de que los datos no sean aportados por las personas interesadas, entonces, el artículo 13 del RGPD señala que dicha información se suministrará dentro de un plazo razonable a contar desde el momento en que se obtengan los datos, con un máximo de un mes. Si estos datos deben utilizarse para comunicarse con la persona interesada, en la propia comunicación, habrá que enviarle la información.

[5] HERNÁNDEZ CORCHETE, Juan Antonio. "Transparencia en la información al interesado del tratamiento de sus datos personales y en ejercicio de sus derechos", EN: AA.VV., *Reglamento General de Protección de Datos. Hacia un nuevo modelo europeo de privacidad*, Ed. Reus, Madrid, 2016, p. 207.

[6] AEPD/ APDCAT/ Agencia Vasca de Protección de Datos, *Guía para el cumplimiento del deber de informar*, disponible en: https://www.aepd.es/documento/guia-modelo-clausula-informativa.pdf (consulta:28/08/2024), p. 6.

## 4. La gratuidad de la información

El artículo 12 del RGPD completa esta cuestión, determinando que toda información y comunicación a la persona afectada será a título gratuito a los efectos de evitar que se limite el derecho para aquellas personas que no puedan o no están dispuestas a abonar lo que se les demande.

De esta regla general, se excepcionan a las solicitudes manifiestamente infundadas o excesivas, especialmente debido a su carácter repetitivo. En estos supuestos, el responsable del tratamiento tendrá dos opciones: o bien podrá cobrar un canon razonable en función de los costes administrativos afrontados para facilitar la información, la comunicación o el desarrollo de la actuación solicitada, o bien cumplir la solicitud. Es importante destacar que la persona responsable del tratamiento deberá tomar esta decisión, teniendo en cuenta que recaerá sobre él la carga de demostrar el carácter manifiestamente infundado o excesivo de la solicitud.

## 5. El principio de información por capas

Por último, el RGPD diferencia el contenido de la información que deber suministrar el responsable del tratamiento a la persona afectada según los datos personales hayan sido aportados o no por ella. En ambos casos, se demanda la entrega de una primera información básica, para añadir con posterioridad otra complementaria. Este enfoque se denomina "*información por capas*", que consiste en presentar, en un primer momento, una primera capa que recoja la información básica y resumida, en el mismo momento y por el mismo medio en que se recojan los datos personales y, posteriormente, una segunda, a través de la cual se remite el resto de las informaciones de forma detallada, en un medio más adecuado para su presentación y comprensión[7].

## 6. La información sobre los derechos relacionados con el tratamiento

Otra de las manifestaciones del principio de transparencia es la obligación de informar a las personas afectadas por el tratamiento de sus datos personales, también a las trabajadoras, sobre sus derechos a ejercitar sobre dicho tratamiento. Se trata del derecho de acceso, rectificación y supresión, así como el de portabilidad. En todo caso, aquí solo se va a mencionar a los dos primeros por su vinculación más estrecha con el derecho a la información.

Desde la STC 254/1993, de 20 de julio, se puede deducir que el derecho de acceso está vinculado al derecho a la información, en cuanto a que en la actualidad el RGPD regula en su artículo 15 que la persona interesada podrá ejercer el derecho de acceso

[7] AEPD/ APDCAT/ Agencia Vasca de Protección de Datos, *Guía para el cumplimiento del deber de informar*, DISPONIBLE en: https://www.aepd.es/documento/guia-modelo-clausula-informativa.pdf (consulta:28/08/2024), p. 5.

cuando no ha recibido comunicación previa sobre el tratamiento de sus datos. Así, se reconoce el derecho de las personas interesadas a obtener confirmación de la persona responsable del tratamiento sobre si se están tratando o no sus datos personales y, en su caso, a obtener información sobre dicho tratamiento. Se trata de la misma información que debería haberse dado a las personas trabajadoras cuyos datos están siendo tratados de acuerdo con las pautas ya antes analizadas. De forma que, si se cumpliese con el derecho de información de forma estricta, no tendrá sentido el ejercicio del derecho de acceso.

Por otra parte, el RGPD, en su artículo 16, reconoce el derecho de rectificación sobre los datos personales cuando son inexactos, y a completarlos cuando sean incompletos. El artículo 14 de la LOPD indica que en la solicitud del ejercicio del derecho es preciso establecer a qué datos se refiere y la corrección que habrá que hacerse. Con este objeto, deberá acompañar a la petición, cuando sea preciso, la documentación justificativa de la inexactitud o carácter incompleto de los datos objeto de tratamiento.

## 7. La obligación de informar frente a la sanción en caso de incumplimiento

Será la persona responsable del tratamiento quien debe poder acreditar efectivamente que ha cumplido ese deber de información con el fin de garantizar la licitud del tratamiento. El modo más fácil de mostrar esta evidencia es a través de un documento escrito y firmado, por ejemplo, mediante distintos indicios que lleven inequívocamente a dicha conclusión tales como la recogida con la nómina, en documento como anexo al contrato de trabajo, enviando un e-mail de respuesta a las personas trabajadoras, notificación por correo certificado o a través de la Intranet de la empresa, incluso podría efectuarse situándose la información en tablón de anuncios de la empresa seguro, accesible y visible con facilidad por las personas afectadas[8].

No obstante, dicho todo lo anterior, el artículo 74 de la LOPD considera infracción leve el incumplimiento del ejercicio del derecho a la información o su cumplimiento sin seguir las reglas indicadas en los artículos 13 y 14 del RGPD. Es destacable que, si las sanciones por la falta de información no se consideran infracciones graves y, por tanto, no se sancionan suficientemente, el deber de informar tendrá poca efectividad, dado que a la empresa le podrá resultar más gravoso el cumplimiento de esta obligación que sufrir la sanción por el incumplimiento.

De acuerdo con lo señalado, sería adecuado que se modificase la calificación de las infracciones relativas al derecho a la información con el objeto de conseguir su correcto cumplimiento, elevándolas de leves a graves, sobre todo en los supuestos en los que la información no acompaña, sino que sustituye al consentimiento y es el único instrumento de protección de las personas afectadas por el tratamiento de los datos personales.

[8] *Informe Jurídico de la AEPD*, núm. 0325/2009.

## IV. La protección de datos de carácter personal como límite del derecho a la información colectiva

En este epígrafe, se va a examinar la otra cara de la moneda. Esto es, como la información que tiene que suministrar la empresa a la representación de las personas trabajadoras se ve limitada por el derecho a la protección de datos personales de estas últimas.

La empresa tiene la obligación de informar sobre una serie de cuestiones a la representación de las personas trabajadoras y, como se puede imaginar, en dicho suministro de información se incluye una gran cantidad de datos personales. No obstante, tampoco se exige en este supuesto el consentimiento expreso de las personas trabajadoras afectadas sobre la información que la empresa transmite a su representación, dado que la licitud del tratamiento de los datos personales viene avalada por el ejercicio de derecho de información reconocido en la legislación a dicha representación[9].

En definitiva, solamente cuando se trate de suministrar datos personales de las personas trabajadoras más allá de la obligación legal o convencional de información a la representación que tiene la empresa, entonces será preciso el consentimiento de la persona afectada[10].

A continuación, se van a poner diversos ejemplos que clarifican lo señalado sobre cuándo sí o cuándo no es preciso el consentimiento de las personas trabajadoras para poder recibir la información por su representación.

### 1. Información que no precisa de consentimiento

En primer lugar, la STC142/1993, de 22 de abril, analiza si el salario de la persona trabajadora debía ser preservado de la información dada a la representación. Es decir, si era más adecuado limitar esta información en cuanto a que no aportaba ningún aspecto necesario para el ejercicio de las funciones de dicha representación. Se parte de que la retribución es un dato económico dentro de la relación laboral, que no traspasa la vida personal y, por tanto, puede ser comunicado a la representación de las personas trabajadoras[11].

Asimismo, en la misma línea, se ha entendido que la negativa a suministrar información sobre los datos relativos al complemento de productividad e incentivos de la empresa a la sección sindical vulnera el derecho a la libertad sindical, de forma que es lícita la información, sin el consentimiento de las personas trabajadoras afectadas[12], dado que no se excede de las materias que deben conocer para ejercer sus funciones legales.

Asimismo, se admite el acceso a la información sobre el contenido de los boletines de cotización para la Seguridad Social, donde se refleja la identificación de la empresa,

[9] STSJ de Andalucía, de 5 de octubre de 2010, comentada por APILLUELO MARTÍN, Martín, "Derecho de información del delegado sindical a las retribuciones de los trabajadores y derechos a la libertad sindical y a la protección de datos de carácter personal". *Revista Doctrinal Aranzadi Social*, núm. 28/2011.

[10] *Informe Jurídico de la AEPD*, núm. 0013/2016, p. 5.

[11] STS de 19 de febrero de 2009 (Rº 6/2009, Sala de lo Social).

[12] STSJ del País Vasco, de 16 de mayo de 2006 (Rº 758/2006).

la determinación de la deuda total, y la relación nominal de personas trabajadoras, con independencia de que estos documentos también contengan, además, los datos relativos a la identificación de estas, sus bases de cotización y las prestaciones que les hayan sido satisfechas en régimen de pago delegado[13].

En este contexto, también pueden surgir conflictos en la petición de datos sobre las personas trabajadoras para la organización de las elecciones sindicales. No solo es preciso conocer los nombres de las personas trabajadoras, sino también su edad, su antigüedad y su tipo de contrato para poder calcular el número de personas representantes que deben elegirse o saber quién tiene derecho a votar. Como estos datos personales son precisos para el desarrollo de este derecho a la representación, en el mismo sentido que lo dicho hasta este momento, habrá que concluir que su cesión a la representación viene avalada por la norma laboral.

## 2. Información que sí precisa de consentimiento previo

Sin embargo, no se llega a la misma conclusión respecto a la información de las nóminas de las personas trabajadoras, puesto que en ellas sí se recogen datos más allá de la mera retribución. No se puede entregar una copia de este documento a la representación de las personas trabajadoras, dado que la inclusión de ciertos datos va más allá de los que precisa para cumplir su función representativa. Lo que no es comprensible es que, sin embargo, sí se haya admitido su acceso sin consentimiento a los boletines de cotización, tal y como ya se ha señalado.

Tampoco queda amparada la información sobre el domicilio particular de las personas trabajadoras para enviar información sobre la acción sindical. Como hay otros medios para informar a las personas trabajadoras, no se entiende preciso autorizar esta información a los efectos de ejercer las funciones reconocidas a la representación y, por esto, en este caso, sí es preciso el consentimiento de las personas trabajadoras para acceder a ellos[14].

La misma solución se aplica sobre la trasmisión de información acerca del cuestionario de evaluación de las personas trabajadoras a la representación. Como no queda dentro del contenido de la información que la legislación avala, hay que concluir que solo se puede remitir esta información con el consentimiento de las personas interesadas[15].

## 3. Sobre la entrega de copia del contrato de trabajo

Muy relacionada con esta cuestión, se encuentra la obligación legal de entregar la copia básica del contrato de trabajo a la representación de las personas trabajadoras por parte de la empresa con el fin de que comprobar la licitud de dicha contratación.

[13] *Informes Jurídicos de la AEPD*, núms. 0488/2009 y 0300/2009.

[14] STSJ de Andalucía/Granada, de 18 de julio de 2007 (Rº 1986/2007).

[15] *Informe Jurídico de la AEPD*, núm. 0071/2010.

El problema surge cuando se trata de determinar cuáles son los elementos que deben incluirse en dicha copia para que la representación pueda ejercitar sus labores. Sin duda, no debe constar el número de Documento Nacional de Identidad, ni el domicilio, ni el estado civil, ni cualquier otro dato que pudieran afectar a la intimidad de las personas trabajadoras[16], porque tampoco son necesarios para el ejercicio de sus funciones.

La ya mencionada STC 142/1993 reconoce a las personas trabajadoras, quienes entiendan que el contenido de la copia básica incluye datos personales de carácter reservado, el ejercicio de su derecho de oposición ante el órgano judicial competente a el tratamiento de dichos datos.

## 4. La aplicación del principio de minimización

En este contexto, debe aplicarse especialmente el principio de minimización a tenor del artículo 5 del RGPD, que consiste en que el tratamiento se limite a la finalidad perseguida[17], sin ir más allá de esta. De acuerdo con lo señalado, la empresa debe reducir los datos personales que se facilitan a la representación de las personas trabajadoras. Será preciso que no se suministren aquellos que no sean necesarios para la finalidad perseguida, que no es otra que facilitar a la representación aquellos que sean precisos para defender los intereses de las personas trabajadoras de acuerdo con sus funciones reconocidas legalmente.

En definitiva, la empresa únicamente pondrá a disposición de la representación los datos personales de las personas trabajadoras cuando se precisen para el desarrollo de su derecho legal de información y consulta, pero siempre teniendo en cuenta que no se deben suministrar más datos de los necesarios para el cumplimiento de sus funciones.

Aunque hay que poner en conocimiento de la representación el menor número de datos personales posibles, sin embargo, no puede perderse de vista que, al menos, deben recibir los necesarios para que puedan defender los derechos de las personas trabajadoras y cumplir con sus funciones, especialmente la vigilancia del cumplimiento de la normativa laboral.

Además, recibida la información, hay una doble responsabilidad en la aplicación de las medidas de seguridad en la trasmisión de los datos personales: por una parte, la de la propia empresa, también en relación a los datos suministrados por la representación; pero, igualmente, de esta que debe cuidar que los datos enviados por la empresa no se extravíen o se trasmitan a otras personas[18].

---

[16] *Informe Jurídico de la AEPD*, núm. 0437/2008.

[17] ARTICLE 29 DATA PROTECTION WORKING PARTY, *Opinion 2/2017 on data processing at work*, 8 de junio de 2017, p. 5.

[18] SAN de 15 de julio de 2010 (Rº 560/2009, sala contencioso-administrativa), donde se analiza un caso donde se localiza un fichero en internet, que contenía datos personales de casi 1000 registros de afiliados a sindicatos.

## V. El RGPD, el principio de transparencia y las decisiones automatizadas

En otro orden de cosas, se entiende de interés en este artículo el examen de la incidencia del principio de transparencia en las decisiones automatizadas de las empresas. En concreto, la regulación que desde el Reglamento General de Protección de Datos se recoge sobre este tema. Si bien es verdad que todavía es muy incipiente el uso de la inteligencia artificial en las empresas para organizar las relaciones laborales, se entiende que su utilización va ir creciendo progresivamente y, por tanto, es preciso que ya se vayan dibujando sus límites.

La Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital de 26 de enero de 2022, del Parlamento Europeo, el Consejo y la Comisión, ya advierte que es preciso adoptar el compromiso de velar por la transparencia en el uso de los algoritmos y la inteligencia artificial y, asimismo, añaden la necesidad de garantizar que las personas afectadas estén informadas.

En el ámbito de la empresa, la utilización de decisiones automatizadas dirigidas a organizar las relaciones laborales puede provocar fácilmente vulneraciones de los derechos de las personas trabajadoras[19], debido a los propios sesgos de los algoritmos. En este contexto, es fundamental que los criterios para determinarlos sean públicos y accesibles, de modo que puedan ser consultados por aquellas personas que sean sometidas a los procedimientos, asegurando su transparencia, comprobando especialmente que los criterios utilizados no introduzcan discriminaciones.

Sin entrar en el análisis de los sesgos de los algoritmos y sus consecuencias, cuestión que por sí misma podría ser objeto de otro artículo, no es de extrañar que en el Considerando 57 del 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 (en adelante, RIA) se recoja que la utilización de sistemas de inteligencia artificial en las relaciones laborales tiene un alto riesgo, porque 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. Pero, además, añade, que los sistemas de inteligencia artificial 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 de la intimidad.

En línea similar, la Exposición de Motivos del RGPD pone de manifiesto que "*la protección de las personas físicas debe ser tecnológicamente neutra y no debe depender de las técnicas utilizadas*", añadiendo que la protección de las personas físicas asimismo debe ser extendida "*al tratamiento automatizado de datos personales*". En este ámbito es importante la cautela de la participación de la persona humana en dichos tratamientos.

[19] Se entiende por decisión automatizada, aquella decisión generada automáticamente de un valor a partir de datos personales que es transmitido por la responsable del tratamiento a otra persona, también responsable del tratamiento, y esta última, de un modo determinante, basa una decisión sobre la persona en dicho valor trasmitido. Vid. STJUE de 7 de diciembre de 2023 (asunto C-634/21, Schufa).

## 1. La intervención humana en las decisiones automatizadas

El artículo 22 del RGPD proclama que "*todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar*" [20]. Así, de acuerdo con este tenor, las personas tienen derecho a no ser objeto de una decisión basada exclusivamente en el tratamiento automatizado, siendo precisa la actuación humana para garantizar la protección de las personas afectadas. En concreto, se requiere una intervención humana "*responsable*", donde se aporte capacidad para analizar la información y para, en su caso, modificarla[21], con el objeto de eliminar posibles discriminaciones a través de los sesgos. Habrá que suministrar información sobre los tratamientos y sus procesos a quien sea competente para su valoración en conjunto y tenga autoridad sobre la decisión[22].

No obstante, no todas las decisiones automatizadas deberán ser intervenidas, sino que a tenor del artículo 22 del RGPD, solo lo serán aquellas que produzcan efectos jurídicos o afecten significativamente a las personas de forma similar o parecida a la que puede afectar una decisión jurídica. En el ámbito laboral, por tanto, será necesaria la intervención humana cuando los derechos fundamentales de una persona trabajadora vayan a quedar afectados, pero también cuando puedan influir en otros derechos también considerados esenciales. Por ejemplo, no solamente será precisa cuando pueda quedar vulnerado su derecho fundamental a la intimidad o a la protección de datos, sino también cuando quede afectado, por ejemplo, su derecho a una retribución adecuada.

En general, se entiende que en el ámbito laboral tendrán que casi ser todas intervenidas, excepto las que no sean ejecutivas o sean de mera gestión, puesto que llevarán aparejadas efectos jurídicos o similares para las personas trabajadoras.

La necesidad de implicación humana, en todo caso, se valorará de acuerdo con el impacto que la decisión pueda tener en las relaciones laborales: deben intensificarse o atenuarse dependiendo de las materias sobre las que haya que tomar las decisiones, dado que no es lo mismo su aplicación, por ejemplo, en la determinación de los criterios de selección en los despidos colectivos que en el reparto de tareas cotidianas.

El artículo 22 del RGPD recoge una excepción de la excepción, señalando que se podrá llevar a cabo la decisión automatizada sin intervención humana, cuando exista un

[20] Por su parte, en la misma línea, el artículo 14 del RIA regula la vigilancia humana, señalando que su objetivo es la prevención o reducción al mínimo de los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando un sistema de inteligencia artificial de alto riesgo se utilice 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 otras cautelas.

[21] Grupo de trabajo sobre protección de datos del Art. 29 (GT29), *Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679,* de 3 de octubre de 2017, revisadas el 8 de febrero de 2018, pp. 21- 23.

[22] *Información algorítmica en el ámbito laboral. Guía práctica y herramienta sobre la obligación empresarial de información sobre uso de algoritmos en el ámbito laboral*, Ministerio de Trabajo y Economía Social, mayo 2023, p. 9.

consentimiento explícito de la persona afectada, pero también cuando la decisión sea necesaria para la celebración o la ejecución de un contrato entre la persona interesada y la responsable del tratamiento. Lo cual podría entenderse que es aplicable al ámbito laboral en cuanto a que se puede justificar su uso en la necesidad de celebración o ejecución del contrato de trabajo. No obstante, si esta conclusión se generaliza llevaría a incluir en esta excepcionalidad a todas las decisiones, por una parte, de selección de personal y actos precontractuales y, por otra, a todas las relativas a la ejecución de la actividad laboral, lo que, sin lugar a dudas, ahondaría en el desequilibrio de las partes en la relación de trabajo, al poner al albur de los algoritmos casi cualquier decisión que afectase a la persona trabajadora. Desde mi opinión, no se entiende que se puedan incluir aquí las decisiones que afecten a la ejecución del contrato laboral de forma generalizada, más allá de las que tengan que ver con cuestiones de mera gestión.

En cualquier caso, en estos casos excepcionales, no se puede olvidar que siempre la persona responsable deberá adoptar las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos de las personas interesadas. De esta forma, como mínimo, tendrá la opción de expresar su punto de vista sobre esta decisión e incluso impugnarla[23]. Es decir, al menos, en este contexto, incluso en los casos excepcionados, se conseguirá siempre una intervención humana secundaria a través de estas posibles actuaciones de la persona responsable.

Volviendo al tema general, en este contexto, por tanto, habrá que distinguir entre la utilización de la inteligencia artificial como mero instrumento de ejecución/gestión y su emplea en la toma de decisiones. Y en este supuesto es cuando habrá que establecer cómo se va a imbricarse la inteligencia artificial en el poder de dirección, cuando tome decisiones, sin la intervención directa de la empresa. Puesto que allí es donde habrá que establecer los límites para evitar la vulneración de los derechos fundamentales, y en este contexto el derecho a la información desde sus vertientes individual y colectiva serán elementos intrínsecos para garantizar estos objetivos.

Es interesante señalar que la Directiva del Parlamento Europeo y del Consejo relativa a la mejora de las condiciones laborales en el trabajo en plataformas digitales, en su Considerando 48 y en su artículo 10, determina que hay que tener en cuenta la necesidad de la supervisión humana de los sistemas automatizados, así como de evaluación periódica del impacto que tienen las decisiones, proponiendo incluso la participación de la representación de las personas trabajadoras en dicho proceso. Para cumplir con esta vigilancia se indica que es preciso que las plataformas digitales garanticen los recursos humanos suficientes para la vigilancia de los sistemas automatizados, que deben tener competencia, formación y autoridad para ejercer su función y deben estar protegidas contra consecuencias negativas (como el despido u otras sanciones) en caso de anular las decisiones automatizadas.

[23] Evaluación de la intervención humana en las decisiones automatizadas | AEPD.

## 2. El derecho de la información: la apuesta por extender la regulación de la Directiva de Plataformas a todos los sectores

A medida que se impongan las decisiones automatizadas en las relaciones laborales, el centro del poder de dirección se moverá hacia los procesos automatizados. El equilibrio entre este y los derechos fundamentales de las personas trabajadoras se trasladará incluso a un momento previo al desarrollo de la relación laboral, que puede coincidir con uno anterior a la existencia del contrato entre las partes, donde la empresa estableció las la reglas de desarrollo de sus relaciones laborales, y esto llevará a que el principio de transparencia tenga un lugar aún más privilegiado en este ámbito.

De forma que si las decisiones automatizadas son las que dirigen la relación laboral a partir de unas instrucciones iniciales de la empresa, el derecho de información se convierte en un elemento fundamental para conseguir el equilibrio entre las partes y controlar si su cumplimiento se ajusta a las exigencias del derecho del trabajo.

En este sentido, el Considerando 92 y el artículo 26 del RIA señalan que la regulación contenida en esta norma sobre la inteligencia artificial se entiende sin perjuicio de la obligación de las empresas de informar o de informar y consultar a las personas trabajadoras o a su representación, en virtud del Derecho o las prácticas nacionales o de la Unión, incluida la Directiva 2002/14/CE del Parlamento Europeo y del Consejo, por la que se establece un marco general relativo a la información y a la consulta de los trabajadores, sobre la decisión de poner en servicio o utilizar sistemas de inteligencia artificial. Y añade que es necesario velar por que se informe a las personas trabajadoras y a su representación sobre el despliegue previsto de sistemas de inteligencia artificial de alto riesgo en el lugar de trabajo, incluso aunque no se cumplan las condiciones de las citadas obligaciones de información o de información y consulta previstas en otros instrumentos jurídicos, siendo este derecho de información necesario para garantizar la protección de los derechos fundamentales de las personas trabajadoras.

No obstante, de acuerdo con el artículo 22 del RGPD el derecho individual de información solo se mantiene en el caso de que quede afectada la persona por una decisión automatizada sin intervención humana[24], y no cuando sí lo sea.

En el mismo sentido, los artículos 13.2.f) y 14.2.g) del RGPD configuran los denominados derechos de explicación sobre la lógica aplicada por las decisiones exclusivamente automatizadas[25], que colaboran en que la participación humana, al menos, se consiga a través de información sobre los procedimientos que debe ser de fácil acceso, utilizando formas claras y exhaustivas[26].

---

[24] *Información algorítmica en el ámbito laboral. Guía práctica y herramienta sobre la obligación empresarial de información sobre uso de algoritmos en el ámbito laboral*, Ministerio de Trabajo y Economía Social, mayo 2023, p. 15.

[25] SAEZ LARA, C., "El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación" en *Temas Laborales*, núm. 155/2020, p. 55.

[26] Grupo de trabajo sobre protección de datos del artículo 29, *Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679*, de 3 de octubre de 2017, revisadas el 8 de febrero de 2018, p.28.

Con el fin de conseguir una protección adecuada de las personas trabajadoras en este contexto sería preciso trasladar al ámbito de todos los sectores laborales lo que establece la Directiva del Parlamento europeo y del Consejo relativa a la mejora de las condiciones laborales en el trabajo en plataformas digitales, la cual presenta un doble sistema de transparencia que diferencia entre los sistemas de supervisión (aquel sistema que se utiliza para supervisar, controlar o evaluar, por medios electrónicos, la ejecución del trabajo de personas que realizan trabajo en plataformas o actividades realizadas en el entorno laboral, en particular mediante la recopilación de datos personales, o que sirven para respaldar tales acciones); y los de toma de decisiones (aquel en el que se utilizan para adoptar o respaldar, por medios electrónicos, decisiones que afectan significativamente a personas que realizan trabajo en plataformas, también a las condiciones laborales de trabajadores de plataformas, en particular decisiones que afecten a su contratación, su acceso a las tareas asignadas y a la organización de estas, sus ingresos, en particular la fijación del precio de tareas individuales, su seguridad y su salud, su tiempo de trabajo, su acceso a formación, promoción o equivalente, o a su situación contractual, incluida la restricción, suspensión o cancelación de sus cuenta). En el segundo de los casos se reconoce una mayor garantía a través del principio de transparencia en cuanto a que, dadas sus características, es más posible la vulneración de los derechos de las personas trabajadoras.

El artículo 9 de la Directiva se titula "*Transparencia en sistemas automatizados de supervisión o de toma de decisiones*", y reconoce el derecho a la información de las personas que realizan trabajo en plataformas, a su representación y, si así lo solicitan, a las autoridades nacionales competentes, sobre el uso de sistemas automatizados de supervisión o de toma de decisiones[27].

Pero, a más, en su artículo 11 declara que los Estados miembros velarán por que las personas que realizan trabajo en plataformas tengan derecho a obtener una explicación de la plataforma digital de trabajo, sin demora indebida, en relación con cualquier decisión adoptada o respaldada por un sistema automatizado de toma de decisiones.

En primer lugar, se reconoce la obligación de la plataforma digital de ofrecer a las personas implicadas la posibilidad de debatir y aclarar los hechos, las circunstancias y los motivos de tales decisiones que afectan significativamente a sus condiciones de trabajo, incluso indicando una persona de contacto perteneciente a la plataforma. Esto es, el derecho de información se extiende a los efectos de conseguir que las personas trabajadoras comprendan las posibles afectaciones de sus relaciones laborales.

[27] Se reconocen diferencias de contenido de la información suministrada a las personas que realizan trabajo en plataformas y la enviada a sus representantes. Las personas trabajadoras tienen derecho directo a recibir información concisa, pudiendo acceder también a información exhaustiva, cuando así lo soliciten. Mientras que la representación ya tiene reconocido directamente el derecho a la información exhaustiva. El derecho se extiende incluso a las personas que aún no son trabajadoras de la empresa, pero están inmersas en procedimientos de contratación o selección, con el mismo contenido que ya se ha señalado, pero referida solamente a los sistemas automatizados de supervisión o de toma de decisiones utilizados en dicho procedimiento, estableciendo que se facilitará siempre antes del inicio del procedimiento.

Por otra parte, el artículo exige a las plataformas digitales que justifiquen por escrito cualquier decisión de restringir, suspender o cancelar la cuenta de la persona trabajadora de plataforma, de denegarle la remuneración por el trabajo que ha realizado o de alterar su situación contractual. En estos casos lo que se le reconoce es un derecho a la información justificada, que parece de menor entidad que el un derecho de explicación. No obstante, como se encuentran situados en el mismo precepto, hay que entender que se trata de una prerrogativa a más, es decir, que se presentará el derecho de información con la justificación en estos supuestos, sin que esto signifique que queda excluido del derecho de explicación, que se aplica a todos los sistemas automatizados de decisiones.

Con independencia de esta cuestión, en caso de que la explicación obtenida no sea satisfactoria o cuando las personas trabajadoras de las plataformas consideren que sus derechos han sido vulnerados, este derecho de explicación lleva aparejado la solicitud de la revisión de la decisión y el derecho a obtener una respuesta motivada en el plazo de dos semanas. Y continúa señalando que las plataformas tendrán que rectificar la decisión sin demora o, si esto ya no es posible, proporcionar una indemnización adecuada, si la decisión vulnera los derechos de la persona trabajadora de plataforma.

En España, la Ley 12/2021, de 28 de septiembre, por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de las plataformas, reconoce el derecho de información a la representación de las personas trabajadoras, no obstante, se echa de menos la regulación de un derecho de información individual, como desarrollo del artículo 22 del RGPD, cuando existan decisiones donde han intervenido procedimientos automatizados[28], sobre todo porque si no existe representación, por el momento, no se asegurará la transparencia.

Pero, más allá de la propuesta de la norma española que se circunscribe al concepto de persona trabajadora y un derecho de información básico a su representación, se precisa una extensión del principio de transparencia en la toma de decisiones automatizadas en las plataformas. Por lo que cuando entre en vigor esta Directiva, sin lugar a dudas, será precisa la actualización de los limitados derechos de información de las partes implicadas, que se reconocen en la actualidad.

## VI. La ampliación del principio de transparencia: desde el mero derecho de información al derecho de explicación

Como ya se ha mencionado, el artículo 13.2.f) del RGPD señala que el responsable del tratamiento facilitará a la persona interesada, en el momento en que se obtengan los datos personales, información sobre la existencia de decisiones automatizas, sobre la lógica aplicada, así como sobre la importancia y las consecuencias previstas de dicho

[28] GORELLI HERNÁNDEZ, J., "Algoritmos y transparencia. ¿pueden mentir los números? Los derechos de Información" en *Trabajo y Derecho*, núm. 86/2022.

tratamiento para el interesado a los efectos de garantizar un tratamiento de datos leal y transparente. Por otra parte, el artículo 14.2.g) del mismo cuerpo legal determina que igualmente habrá que suministrar la misma información, cuando esta no se haya obtenido de la persona interesada.

Así, se configuran los derechos de explicación desde la lógica aplicada por las decisiones exclusivamente automatizadas[29], que colaboran en que la participación humana, al menos, se consiga a través de información sobre los procedimientos, la cual debe ser de fácil acceso, utilizando formas claras y exhaustivas.

Como se puede observar, la explicabilidad es un concepto que va más allá de la noción básica de transparencia, del derecho a la mera información[30], donde se consigue que las personas puedan comprender el contenido de dicha información. Y es justamente desde esta comprensión, desde donde se podrán valorar las responsabilidades y las posibles discriminaciones de dichas decisiones[31].

La Recomendación 40 sobre la Ética de la Inteligencia Artificial de la UNESCO, de noviembre de 2021, señala que la explicabilidad supone hacer inteligibles los resultados de los sistemas de inteligencia artificial y facilitar información sobre ellos. Se refiere a la inteligibilidad de la entrada, salida y funcionamiento de cada componente algorítmico y la forma en que contribuye a los resultados de los sistemas. Así pues, la explicabilidad está estrechamente relacionada con la transparencia, en cuanto a que los resultados y los subprocesos que conducen a ellos podrán ser comprensibles y trazables.

Y después de relacionar explicabilidad con transparencia señala que quien sea responsable del algoritmo debería comprometerse a que sea explicable, sobre todo cuando su impacto no sea temporal y fácilmente reversible o de bajo riesgo. Y, como de acuerdo con el RIA, la utilización de la inteligencia artificial en el ámbito de las relaciones laborales se califica de alto riesgo, habrá que deducir que la explicabilidad es un elemento que debe acompañar siempre al derecho de la información cuando se trata del ámbito de las decisiones automatizadas.

El derecho explicación es un paso más en la configuración del principio de transparencia, superando el derecho a la información y consiguiendo una mayor intensidad en la protección. Dicho esto, si en algunas situaciones parece adecuado la elección de esta graduación en la escala de la transparencia, en el contexto de los procesos automatizados, es imprescindible.

No obstante, dada la complejidad de la comprensión general del tratamiento de los datos personales y también, en concreto en el ámbito de las relaciones laborales, quizás el principio de información, al menos, el individual de las personas afectadas, debería

[29] SAEZ LARA, C., "El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación" en *Temas Laborales*, núm. 155/2020, p. 55.

[30] NIST, AI Risk Management Framework: Initial Draft, 17 de marzo de 2022, pp. 13 y ss. (AI Risk Management Framework: Initial Draft - March 17, 2022 (nist.gov)

[31] TODOLI SIGNES, A., "El principio de transparencia algorítmica en su dimensión individual y colectiva: especial referencia a la Directiva de Plataformas Digitales y al Reglamento de IA" en *Trabajo y Derecho*, núm. 19/2024, p. 4.

transitar al derecho de explicación de forma genérica, con el fin de asegurar la protección adecuada y la comprensión de sus derechos frente a estos tratamientos.

En este sentido, hay que recordar que el artículo 88 del RGPD reconoce la posibilidad de ampliar la protección de los derechos en relación a la protección de datos personales de las personas trabajadoras en el ámbito de las relaciones laborales con el fin de, entre otros, de garantizar la transparencia del tratamiento. Además, esta obligación podría estar avalada igualmente por el artículo 12 del mismo cuerpo normativo, que declara que la información sobre el tratamiento a las personas afectadas debe ser comprensible y entendible, lo que lleva quizás a que no pueda entenderse cumplida con el mero derecho a la información, sino que precisa, en cierto modo, de este derecho de explicabilidad, no solo en relación a las decisiones automatizadas, sino a todo tratamiento que las afecte.

# La regulación de la inteligencia artificial en la Directiva de trabajo en plataformas digitales*

## The regulation of artificial intelligence in the work Directive on digital platforms

Adrián Todolí Signes**
*Prof. Titular de Derecho del Trabajo y de la Seguridad Social*
*Universidad de Valencia*
ORCID ID: 0000-0001-7538-4764

doi: 10.20318/labos.2024.9030

*Resumen:* La recientemente aprobada Directiva de trabajo en plataformas digitales contiene dos partes bien diferenciadas. De un lado, regula una presunción de laboralidad para todos los trabajadores de plataformas digitales. De otro lado, establece garantías, individuales y colectivas, para trabajadores y autónomos frente a la dirección algorítmica (o uso de la Inteligencia Artificial en el trabajo). El objeto de este trabajo es el análisis de la segunda parte de la normativa: las protecciones frente al uso de la inteligencia artificial, para dirigir y controlar el trabajo, existentes en la Directiva. Para ello, en primer lugar, se analiza el ámbito de aplicación de la Directiva dado que el concepto de "plataforma digital" requiere ser interpretado para conocer a qué plataformas les resulta de aplicación la normativa. En segundo lugar, respecto a las garantías frente a la dirección algorítmica se estudia su relación con otras garantías existentes anteriormente como es el RGPD, así como el alcance contenido y problemas interpretativos que surgen de la nueva normativa. Por último, resultado del análisis y de los problemas interpretativos encontrados se realiza una serie de recomendaciones para su transposición.

*Palabras clave:* Directiva de trabajo en plataformas digitales, DTPD, trabajo en plataformas, uber, dirección algorítmica del trabajo, Inteligencia artificial, concepto de plataforma digital de trabajo.

* Investigación que forma parte del proyecto de investigación del Ministerio de Ciencia e Innovación titulado "Algoritmos extractivos y neuroderechos. Retos regulatorios de la digitalización del trabajo" ref. PID2022-139967NB-I00. Adicionalmente, este artículo es el resultado de la ponencia realizada en Congreso denominado La IA en el mundo del trabajo que se desarrolló en la Facultad de Derecho de la Universitat de València los días 20 y 21 de junio.

www.adriantodoli.com

*Abstract:* The recently approved Digital Platform Work Directive consists of two well-differentiated parts. On one hand, it establishes a presumption of employment for all digital platform workers. On the other hand, it sets forth both individual and collective safeguards for workers and self-employed individuals against algorithmic management (or the use of Artificial Intelligence in the workplace). The purpose of this paper is to analyze the second part of the Directive: the protections against the use of artificial intelligence to manage and control work as outlined in the Directive. To this end, the scope of the Directive is first examined, as the concept of "digital platform" requires interpretation to determine which platforms fall under its provisions. Secondly, regarding the safeguards against algorithmic management, the paper explores their relationship with pre-existing protections, such as the GDPR, as well as the scope, content, and interpretative challenges posed by the new regulation. Finally, based on the analysis and the interpretative issues identified, a series of recommendations for its transposition are presented.

*Keywords:* Digital Platform Work Directive, DPWD, platform work, Uber, algorithmic management of work, artificial intelligence, concept of digital work platform.

## 1. Introducción

El trabajo en plataformas digitales ha supuesto en la última década una oportunidad de trabajo para muchas personas. Los estudios plantean que es una alternativa para generar ingresos especialmente para personas inmigrantes y otros colectivos con dificultades para el acceso a un trabajo tradicional[1]. Sin embargo, los análisis también han señalado que el trabajo en plataformas digitales presenta riesgos de precariedad, falta de previsibilidad de ingresos, exceso de horas de trabajo y riesgos laborales específicos[2]. A su vez, la tendencia de las plataformas de clasificar como colaboradores independientes a los prestadores de servicios ha supuesto un reto para la protección social de los mismos. Tribunales en toda Europa han comenzado a reclasificar a los trabajadores de plataformas, en sectores del transporte y reparto, como asalariados tras largos procesos judiciales con elevado coste para los demandantes y los Estados[3].

Con el objetivo de aportar seguridad jurídica, reducir la conflictividad, aminorar la carga probatoria de quién es realmente asalariado y, en general, mejorar las condiciones de trabajo de aquellos que obtienen ingresos en plataformas digitales la UE ha aprobado la Directiva europea de mejora de las condiciones laborales en el trabajo en plataformas (en adelante, DTPD).

En materia de Inteligencia artificial, la norma se dedica profusamente a regular la dirección algorítmica del trabajo en plataformas digitales. Se establecen prohibiciones de procesar datos y derechos de información, individuales y colectivos. Una de las mayores

[1] ILO, *World Employment and Social Outlook The role of digital labour platforms in transforming the world of work*, Geneve, 2021.

[2] Taylor M., et al., *Good work: the Taylor review of modern working practices*, London, 2017. Todolí A (Dir.) *Riesgos específicos de Riesgos laborales específicos del trabajo en plataforma digitales*, OSALAN, 2020.

[3] Hießl, C., "The Classification of Platform Workers in Case Law: A Cross-European Comparative Analysis". *Comparative Labor Law & Policy Journal*, V. 42.2, 2022.

novedades es la obligación de las plataformas de realizar una auditoría algorítmica para garantizar que el algoritmo no está vulnerando derechos de los prestadores de servicios. La cuestión que surge es quién debe realizar dicha auditoría ¿puede ser externa a la empresa? y cuál debe ser su contenido mínimo. Adicionalmente, estos derechos surgidos de la Directiva, incluidos algunos colectivos, no se van a aplicar solo a asalariados, sino también a prestadores de servicios autónomos, sin que la normativa determine de forma expresa quienes pueden tener la consideración de representantes de estos colectivos.

Este artículo tiene por objetivo analizar el contenido de la DTPD en materia de Inteligencia artificial y resolver las dudas jurídicas que plantea, así como, realizar propuestas para la trasposición que limiten la inseguridad jurídica y permitan una mayor efectividad de las garantías concedidas por la norma europea. El artículo continua con el análisis del ámbito de aplicación de la norma y sus posibles efectos sobre agencias de colocación online, plataformas de contenido y de concursos. El epígrafe tercero analiza las garantías frente a la dirección algorítmica. Por último, el trabajo finaliza con recomendaciones para su futura transposición.

## 2. Concepto legal de plataforma aplicado a la variedad tipológica. La dificultad de poner puertas al campo

### *2.1. El concepto de "plataforma digital de trabajo" como ámbito de aplicación de la Directiva*

Como ya se ha señalado, la DTPD se aplica, no solamente a asalariados, sino también a autónomos. De esta forma, el ámbito de aplicación clásico de la normativa laboral que consiste en definir quién es un asalariado y a este se le aplicará la normativa laboral (sin definir quién es el empresario más allá de entender que lo será todo aquél que contrate con un trabajador) no era de utilidad para las pretensiones de la Directiva.

Por el contrario, el ámbito de aplicación de la normativa pasará a ser "el trabajo en una plataforma digital". Esto es, el foco se pone precisamente en el "tipo" de empresa, siendo al empresario de "plataformas digitales" a aquél que, en su relación con el prestador de servicios –sea asalariado o autónomo verdadero–, se le imponen una serie de obligaciones.

El reto de definir qué es una plataforma digital de trabajo es mayúsculo. Desde los primeros trabajos doctrinales en materia de trabajo en plataformas digitales[4] se ha señalado la gran variedad y heterogeneidad que existe de plataformas digitales. Así, solamente por nombrar alguna de las categorizaciones se encontraría: las plataformas digitales de trabajo presencial (Uber); plataformas digitales de trabajo online (iStockphoto); plataformas de *microworking* (Amazon Mecanichal Turk); plataformas de *freelancing* (Fiveer); plataformas digitales de contenidos (Youtube, twitch); plataformas de trabajo genéricas

[4] De Stefano, V., "The rise of the «just-in-time workforce»: On-demand work, crowdwork and labour protection in the «gig-economy", ILO, 2016. Todolí Signes, A., *La era de la economía colaborativa*, Tirant lo Blanch, 2017.

(Field Agent); específicas (Glovo); basadas en concursos (99 designs)[5], etc. A estas cabría sumar las plataformas digitales que publicitan ofertas de trabajo (InfoJobs) o de intermediación en el mercado de trabajo (agencias de colocación digitales).

Adicionalmente, se encuentran otro tipo de plataformas que podrían definirse por el hecho de que la prestación de servicios no es el centro del negocio jurídico subyacente. Aquí se pueden encontrar plataformas de alquiler de bienes (Airbnb, Parquo), de venta de bienes (Wallapop, Amazon); plataformas de reparto de gastos en transporte (Bla Bla Car), etc. Por ello, el reto de definir el "trabajo en plataformas digitales" no solamente consiste en conceptualizar y describir qué es una plataforma digital, sino cuando la misma tiene como actividad la prestación de servicios, descartando aquellas plataformas cuya actividad sea otro negocio jurídico (ej. alquiler).

Con esto en mente, la normativa, es su art. 2.1 DTPD, define las plataformas digitales de trabajo como ámbito de aplicación de la normativa con cuatro requisitos que deben cumplirse acumulativamente.

a) Provee servicios a través de medios electrónicos como una web o una aplicación;
b) ese servició se provee a petición del destinatario del servicio;
c) la plataforma organiza el trabajo realizado por individuos a cambio de una contraprestación económica, el servicio puede desarrollarse online o presencialmente;
d) existen sistemas automatizados de seguimiento o de toma de decisiones

Con esta definición, existe un conjunto de plataformas digitales de trabajo que claramente están incluidas, a saber, las plataformas digitales de trabajo presencial (Uber); plataformas digitales de trabajo online (iStockphoto); plataformas de microworking (Amazon Mecanichal Turk); plataformas de freelancing (Fiveer)[6]. Todo ello con independencia de que sean plataformas dedicadas a un sector de actividad específico o sean plataformas donde, genéricamente, se pueda encontrar trabajadores/autónomos pertenecientes a cualquier sector de actividad.

A su vez, la definición excluye de forma bastante clara cierto tipo de plataformas. En ese sentido, dado que se exige que la plataforma *organice el trabajo* realizado por individuos, esto parece excluir las plataformas cuya actividad subyacente no sea la prestación de servicios. Esto es, se excluiría así plataformas de alquiler de bienes (Airbnb, Parquo), de venta de bienes (Wallapop, Amazon); plataformas de reparto de gastos en transporte

[5] En estas plataformas un cliente ofrece un encargo de abierta a toda una comunidad de personas en la plataforma. Los trabajadores realizan el trabajo y lo entregan y el cliente elige entre todos los trabajos realizados cuál prefiere. Este trabajo elegido es el único que percibe una retribución por el trabajo, ver Todolí Signes, A., *El trabajo en la era de la economía colaborativa*, Tirant lo blanch 2017; Araujo R., "99designs: An Analysis of Creative Competition in Crowdsourced Design", First AAAI Conference on Human Computation and Crowdsourcing, v1, 2013.

[6] Silberman M. S., "The concept of the "digital labor platform", London College of political Technology working paper, 2023, p 2.

(Bla Bla Car). En este sentido, también parecen excluidas las plataformas que solamente se dediquen a ofrecer anuncios de ofertas de trabajo o de prestación de servicios sin intervenir en la selección del concreto prestador de servicios. En apoyo a esta interpratación, el considerando 20 de la DTPD indica que no entran en el concepto de plataforma digital las plataformas digitales que no organizan el trabajo, sino solamente facilitan a los proveedores de servicio encontrar clientes sin más.

### *2.2. Las agencias de colocación digitales*

La definición legal de trabajo en plataformas basada en estos cuatro requisitos aclara la inclusión y exclusión de algunos grandes tipos de plataformas digitales. Sin embargo, existen otros donde las dudas son mayores: por ejemplo, en las agencias de colocación digitales.

En un primer momento se debería rechazar la inclusión de una agencia de colocación en el ámbito de aplicación de la Directiva, dado que no cumple el requisito (c): Una agencia de colocación pura no va a organizar el trabajo ni tampoco a retribuirlo (tras encontrar la persona adecuada para el trabajo es el cliente final el que organiza y retribuye el trabajo). La controversia surge debido a que el considerando 20 de la DTPD establece que organizar el trabajo "debe implicar, como mínimo, un papel importante en vincular la demanda del servicio a la oferta de mano de obra de una persona física (…) puede incluir otras actividades, como la tramitación de los pagos". De esta forma, la DTPD usa un concepto muy amplio de "organización del trabajo". Esto parece coherente con el objetivo de la Directiva de incluir en el ámbito de aplicación aquellas plataformas que cuentan con verdaderos autónomos realizando prestaciones de servicios. Una interpretación de concepto de "organización del trabajo" próxima a los conceptos clásicos de "dirección" o "coordinación" del trabajo –esto es, dependencia– excluiría cualquier plataforma que provea servicios a prestadores de servicios verdaderamente autónomos.

De esta forma, la amplitud en la que el considerando 20 define lo que es "organizar trabajo", a efectos de la Directiva, hace pensar que una agencia de colocación online cumple con el tercero (c) de los requisitos.

Ahora bien, para incluir las agencias de colocación online en el ámbito de aplicación de la DTPD es necesario que se cumpla también el cuarto requisito: el uso por parte de la plataforma, de sistemas automatizados de seguimiento o de toma de decisiones. Ambos sistemas son definidos por el propio texto de la norma. El art. 2.1 (h) establece que los sistemas de seguimiento son aquellos usados "para realizar el seguimiento, controlar o evaluar, por medios electrónicos, la realización del trabajo de personas que realizan trabajo en plataformas o las actividades realizadas en el entorno de trabajo, también mediante la recopilación de datos personales". Por su parte, los sistemas automatizados de toma de decisiones son definidos en el art. 2.1 (i) como "sistemas que se utilicen para adoptar o respaldar, por medios electrónicos, decisiones que afecten significativamente a personas que realicen trabajo en plataformas, también a las condiciones laborales de trabajadores de plataformas, en particular decisiones que afecten a su contratación, su

acceso a las tareas asignadas y a la organización de estas, sus ingresos, incluida la fijación del precio de tareas individuales asignadas, su seguridad y su salud, su tiempo de trabajo, su acceso a formación, su promoción o equivalente, y a su situación contractual incluida la restricción, suspensión o cancelación de sus cuentas".

Este concepto, que se acaba de ver, de "sistema automatizado de toma de decisiones" implicará que solamente podrán estar incluidas en el concepto de plataforma digital de trabajo aquellas agencias de colocación digitales que usen sistemas electrónicos para tomar o apoyar decisiones de "selección" de los concretos trabajadores ofrecidos al cliente. Si, adicionalmente, las "agencias de colocación" recogen evaluaciones (*feedback*) de los clientes respecto al trabajo realizado –como pueda ser la conocida reputación digital de los prestadores de servicios[7]– quedaría aún más claro que encajan en el cuarto requisito. Esto es, de acuerdo con el 2.1 (h), si la agencia de colocación digital usa algoritmos para realizar la casación entre oferta y demanda laboral se podría estar cumpliendo el cuarto requisito (d) del concepto de "Plataformas digitales de trabajo".

Incluso más claramente podrían estar incluidas en estos requisitos las "agencias de colocación digitales" que se dedican únicamente a un tipo de trabajadores. En efecto, una agencia de colocación clásica suele implicar una empresa que casa oferta y demanda de trabajo en todo tipo de sectores según las necesidades de la empresa cliente. Sin embargo, en el mundo digital –derivado de los menores costes fijos que implica el mundo onlin–, muchas de estas plataformas se han especializado en buscar (casar oferta y demanda) proveedores de servicios de una actividad concreta. Así, por ejemplo, son muy conocidas las plataformas de cuidados o de limpieza del hogar en las que la plataforma selecciona mediante su algoritmo el concreto prestador de servicios para luego ser contratado directamente por el cliente final. También ocurre con plataformas de abogados en las que la plataforma selecciona algorítmicamente al abogado concreto que realizará los servicios. O incluso plataformas donde el cliente final recibe varios presupuestos distintos realizados por los contratistas o proveedores de servicios de la plataforma para que el cliente final elija que prestador prefiere. En todos estos casos, en mayor o menor medida, la plataforma está algorítmicamente cruzando oferta y demanda siendo los criterios del algoritmo los que permiten la selección concreta de quién tiene acceso al cliente final de la plataforma y, por tanto, el algoritmo es crucial para determinar la capacidad de obtener ingresos del trabajador o proveedor de servicios autónomo.

Cuestión distinta será el caso en que sea el cliente final, sin intervención de la plataforma, el que elija libremente y sin información o datos provistos por la plataforma, sin elaboración de perfiles, ni monitorización en sentido del art. 2.1 (h) o reputación digital. En estos casos, la plataforma podría quedar excluida del ámbito de aplicación de la DTPD.

En cualquier caso, la gran variedad y tipología de plataformas y sus diversas formas de funcionar hace que, en muchos casos de nuevo, exista una amplia zona gris de incertidumbre sobre la aplicación o no de la Directiva de plataformas.

[7] Todolí Signes, A., The evaluation of workers by customers as a method of control and monitoring in firms: Digital reputation and the European Union's General Data Protection Regulation, *International Labour Review*, 160 (1), 2020, https://doi.org/10.1111/ilr.12161.

### *2.3. Plataformas de creadores de contenidos*

En plataformas como Youtube, Twitch, Onlyfans, etc, miles de personas obtienen ingresos creando contenido que es ofrecido en la plataforma o que es visualizado en tiempo real por los aficionados. Las plataformas insertan publicidad en los videos y pagan a los creadores del contenido una parte de lo obtenido. La cantidad percibida varía según el número de visualizaciones. A su vez, ambos dependen, en gran medida, del diseño realizado del algoritmo de la plataforma. Los algoritmos deciden los videos "recomendados" para la audiencia y la probabilidad de que un usuario vea un video está íntimamente relacionada con que este sea recomendado[8]. También los creadores de contenido obtienen ingresos gracias a los "suscriptores", esto es, aficionados a un determinado "influencer" que pagan mensualmente, a través de la plataforma digital, una cantidad para poder visualizar los videos de su "influencer" favorito o para obtener acceso a alguna funcionalidad extra (como poder escribir al "influencer" alguna petición, tener acceso a contenido exclusivo, etc.)[9].

El funcionamiento de las plataformas de contenido parece incluido en el "trabajo en plataformas digitales" desde una doble perspectiva interpretativa: literal y finalista. Empezando por la interpretación finalista, la norma tiene por objetivo, por un lado, aclarar cuándo se está ante un asalariado y cuando ante un verdadero autónomo y, por otro lado, otorgar derechos frente al uso de algoritmos que afectan a las condiciones laborales tanto asalariados como autónomos.

En este sentido, ambos objetivos son necesarios en las plataformas de contenidos. Por un lado, la doctrina ha sugerido que, si bien las notas de laboralidad no se encuentren tan claramente en plataformas de contenidos como en otro tipo de plataformas, la realidad es que el alto control algorítmico existente respecto al contenido, duración y frecuencia de los videos realizados y las extensas jornadas de trabajo imprescindibles para obtener unos ingresos dignos hace necesario aclarar su estatus jurídico[10]. A su vez, y especialmente, la aplicación de las protecciones frente al uso de algoritmos es necesaria para los creadores de contenido –aunque estos sean verdaderos autónomos–. El diseño del algoritmo y sus modificaciones afecta a la cantidad de suscriptores, de visualizaciones, a sus ingresos, etc. También el diseño del algoritmo modifica la duración de los videos (premiando más unos videos que otros) y su contenido. De esta forma, los derechos de transparencia y de explicación frente a los algoritmos son fundamentales para este tipo de prestadores de servicios.

Respecto a la interpretación literal del concepto, el uso algorítmico y la organización del trabajo en el sentido de casar los "clientes" con el prestador de servicios son claros (requi-

[8] Entre miles de millones de videos existentes en la plataforma, aquellos que son recomendados por el algoritmo son los únicos que tienen una oportunidad real de generar ingresos.

[9] Para ver en profundidad el funcionamiento, ver Pătraș L., Todolí Signes, A., "Ser influencer hoy: posibilidades y obstáculos de una nueva fuente de empleo", 2022 https://www.uv.es/catedra-economia-colaborativa-transformacion-digital/es/novedades/informe-ser-influencer-hoy-posibilidades-obstaculos-nueva-fuente-empleo-1286057015758/Novetat.html?id=1286256097235

[10] Pătraș L., Todolí Signes, A., "Ser influencer hoy: posibilidades y obstáculos de una nueva fuente de empleo", 2022.

sitos (a), (c) y (d)). Ahora bien, el requisito b) ("ese servició se provee a petición del receptor") puede tener mayor dificultad de encaje en las plataformas de contenidos. En efecto, los creadores de contenido normalmente no desarrollan contenido personalizado ni bajo petición específica, sino que producen los videos para toda su audiencia. De esta forma, se aleja del concepto de "petición" o "trabajo bajo demanda" de otro tipo de plataformas.

Así, el cumplimiento del requisito (b) dependerá de cómo se interprete el requisito "petición del receptor". De un lado, puede interpretarse, en sentido estricto, que solamente se cumple el requisito en aquellos casos en los que el servicio prestado se adecua a las necesidades previamente manifestadas de cada cliente.

De otro lado, en una interpretación más amplia, puede entenderse cumplida cuando exista una petición genérica del servicio. En este sentido, el modelo de "suscriptores" implica que los receptores del servicio –la audiencia o los aficionados– están pagando por tener más contenido de ese prestador de servicios. Por ello, esta "petición" puede interpretarse en un sentido amplio, que ese pago de suscripción representa una petición para que ese prestador de servicios (el "influencer") cree más contenido –esto es, un encargo para ver más contenido originado por la persona–. En mi opinión, cuando la norma exige la "petición del receptor", solamente está describiendo la relación triangular clásica de las plataformas digitales. Esto es, la existencia de una plataforma que se interpone entre el cliente o usuario final y el prestador de servicios, sin que parezca que la norma prescriba que la petición deba ser concreta o específica.

En cualquier caso, incluso en los modelos económicos que no funcionan con suscripciones, sino que es la propia plataforma la que paga al "influencer" una parte de lo obtenido de la publicidad, se podría argumentar que quién estaría realizando una petición de encargo de creación de contenido sería la propia plataforma. La demanda de trabajo viene de la plataforma que necesita el contenido para poder insertar la publicidad en él. Esto es, desde el momento en que la plataforma anuncia que pagará a los influencers con mayor audiencia, está solicitando la creación de contenido que tendrá retribución variable dependiendo de la calidad y el éxito de este. Por lo que se cumpliría con el requisito (b). En apoyo a esta segunda interpretación, cabe incluir que el considerando 19 de la DTPD, que contempla la posibilidad de que dentro del concepto de "plataforma digital de trabajo" el receptor del encargo sea la propia plataforma digital.

### *2.4. Plataformas basadas en concursos*

Las plataformas basadas en concursos son aquellas en las que un cliente solicita a los prestadores de servicios que participen en el desarrollo de una tarea (ej, el diseño de un logo, de una página web, de un proyecto de arquitectura) y la envíen. El cliente, entre todos aquellos que desean participar en el concurso, selecciona la ganadora y la retribuye. El resto de los participantes no seleccionados no perciben retribución. Este tipo de plataformas plantean dudas respecto a la aplicación de la DTPD en cuanto a los requisitos (b) y (d), esto es, (b) se considera que la prestación del servicio se realiza a

cambio de un pago; (d) existen medios algorítmicos o automatizados de monitorización o selección.

En cuanto al requisito del pago, pocas dudas presenta la aplicación de la Directiva al ganador del concurso. La duda realmente surge concerniente al resto de usuarios que no son retribuidos por no ser seleccionados en el concurso. En mi opinión, la necesidad de pago debe ser interpretado en el sentido clásico de la retribución en Derecho del trabajo, esto es, como exclusión del trabajo voluntario. Así, lo importante no es si finalmente hay o no pago, sino si existe "ánimo de lucro" al realizar la prestación de servicios, como contraposición del trabajo amistoso o benevolente que sería el trabajo excluido. De esta forma, cuando se presta un servicio con el objetivo de percibir una retribución –aunque finalmente esta no se obtenga– se cumpliría el requisito de "a cambio de un pago" exigido por la DTPD.

El segundo requisito controvertido de esta tipología de plataformas es conocer si se cumple la exigencia de seguimiento o toma automatizada de decisiones (d). En este caso, habrá que estar al caso concreto. En términos generales puede afirmarse que estos requisitos se encontrarían en aquellos supuestos en los que el algoritmo decide quién recibe la oportunidad de participar en el concurso (ej., solo reciben la oferta de participación en el concurso aquellos con una mínima reputación digital), en contraposición con plataformas en las que todos los inscritos reciben indiscriminadamente todas las ofertas de participación en concursos.

También se cumpliría este requisito si la plataforma estableciera procedimientos de "selección" del ganador del concurso, aunque fueran preliminares o parciales, que fueran automatizados. Así, por ejemplo, se cumpliría el requisito si existiera un algoritmo que evaluara los proyectos enviados, pero también si el algoritmo comprobara el cumplimiento de unos requisitos mínimos del proyecto o del usuario para pasar a la siguiente fase de selección del concurso.

### *2.5. El uso de empresas intermediarias entre la plataforma y los prestadores de servicios*

De la experiencia española con la Ley Rider, la DTPD advierte la posibilidad de que las empresas de plataformas usen otras empresas –empresas intermedias– para que contraten y retribuyan el trabajo. Así, la DTPD regula en su art. 3 las empresas intermediarias.

Las empresas intermediarias, a efectos de la DTPD, son aquellas que gestionan la mano de obra que presta el servicio para los clientes de la plataforma. Así, para el servicio de reparto de comida a domicilio, empresas como Ubereats o JustEat contratan a terceras empresas para que realicen los encargos que el cliente final solicita mediante su aplicación. Lo mismo ocurre en el servicio de trasporte por ciudad en el caso de Uber y Cabify. Con esta externalización de la prestación de servicios la plataforma digital se aleja jurídicamente de las responsabilidades de la gestión de la mano de obra que presta el servicio.

Con objeto de evitar que el uso de empresas intermedias entre la plataforma digital y la persona que presta el servicio impida la aplicación de la DTPD, el art. 3 de la

misma regula este supuesto en el sentido de que los EM deben garantizar que las personas prestadoras de servicio que tengan una relación contractual con una intermediaria deben disfrutar del mismo nivel de protección concedido por la Directiva que aquellos que tienen una relación contractual directa con la plataforma.

El objetivo de la normativa parece ser evitar que no se aplique la DTPD por el mero uso de la intermediaria. Sin embargo, la redacción del art. 3 está realizado en términos muy vagos y generales lo que permite varias interpretaciones.

En efecto, la normativa solamente exige que los prestadores "disfruten el mismo nivel de protección" que si la relación contractual fuera directamente con la plataforma digital. La primera pregunta que se plantea es si los derechos de trasparencia y de información individual y colectivos, así como el resto de los derechos en materia algorítmica, podrán ejercerse frente a la intermediaria, frente a la plataforma o frente a los dos.

En mi opinión, una interpretación finalista del art. 3 DTPD debe llevar a entender que los derechos deben garantizarse siempre frente ambos. En efecto, el objetivo del art. 3 DTPD es que el trabajador no sufra perjuicio por la existencia de la intermediaria. De esta forma, si los derechos se ejercen solamente frente a la intermediaria, se encontrarían en peor situación que otros prestadores de servicios comparables cuya relación directa fuera la plataforma digital[11].

## 3. Las protecciones frente a los algoritmos

La DTPD dedica su capítulo tercero a la dirección algorítmica. En la literatura se ha expuesto de forma profusa los riesgos para las condiciones laborales, para los colectivos históricamente más discriminados, para la salud de los trabajadores e incluso para la libertad sindical y el poder de negociación de las personas trabajadoras el hecho de que los algoritmos tomen decisiones que les afecten[12]. En respuesta a estos retos, la DTPD plantea un conjunto de medidas detalladas y completas que otorgan protecciones individuales y colectivas frente al uso de algoritmos (que monitorizan o que toman decisiones) por parte de las plataformas digitales de trabajo.

La normativa tiene por objeto reducir el impacto negativo en las condiciones laborales de esta nueva innovación tecnológica cuando se usa para dirigir, organizar y controlar el trabajo. La DTPD diseña cuatro tipos de protecciones: i) prohíbe cierto tipo de usos de la tecnología; ii) fija obligaciones de trasparencia sobre el uso y funcionamiento de este tipo de tecnología. Los derechos de información son de carácter individual y co-

[11] Para ver una justificación de la "peor" condición en caso de que exista una empresa intermedia ver, Esteve Segarra A., Todolí Signes, A., "Cesión ilegal de trabajadores y subcontratación en las empresas de plataforma digitales", RDS, 95, 2021, 37-64; Todolí Signes, A., "La dirección algorítmica de las redes empresariales: plataformas digitales, inteligencia artificial y descentralización productiva", Revista CEF, 476, 2023.

[12] Ver en extenso Todolí Signes A., *Algoritmos productivos y extractivos, Cómo regular la digitalización para mejorar el empleo e incentivar la innovación*, Aranzadi, 2023.

lectivo; iii) obligaciones de auditoría y control sobre los sistemas algorítmicos; iv) otorga derechos de revisión de las decisiones algorítmicas.

### *3.1. Usos prohibidos*

El capítulo dedicado a la dirección algorítmica comienza limitando el procesamiento de datos personales mediante monitorización automatizada y decisiones automatizadas. De esta forma, el foco de la prohibición se pone en la conjunción de dos elementos, de un lado, procesar datos personales, y de otro lado, con cierto objetivo.

Los objetivos para los que se prohíbe procesar datos personales son los siguientes: (a) procesar cualquier dato personal sobre el estado emocional o psicológico de la persona que realiza trabajo en plataforma; (b) procesar cualquier dato personal en relación con conversaciones privadas, incluidos intercambios con otras personas que realizan trabajo en plataforma y sus representantes; (c) recopilar cualquier dato personal mientras la persona que realiza trabajo en plataforma no esté ofreciendo o realizando trabajo en plataforma; (d) procesar datos personales para predecir el ejercicio de derechos fundamentales, incluidos el derecho de asociación, el derecho a la negociación colectiva y la acción o el derecho a la información y consulta, según se define en la Carta; (e) procesar cualquier dato personal para inferir origen racial o étnico, estado migratorio, opiniones políticas, creencias religiosas o filosóficas, discapacidad, estado de salud, incluida enfermedad crónica o estado de VIH, el estado emocional o psicológico, pertenencia a un sindicato, la vida sexual o la orientación sexual de una persona; (f) procesar cualquier dato biométrico, según se define en el Artículo 4, punto (14), del Reglamento (UE) 2016/679, de una persona que realiza trabajo en plataforma para establecer la identidad de esa persona comparando esos datos con datos biométricos almacenados de individuos en una base de datos.

Como se puede observar para aplicar la prohibición no se exige el uso de un tipo de tecnología algorítmica avanzado ni existen requisitos respecto al tipo de algoritmos o inteligencia artificial. Por el contrario, cualquier tipo de procesamiento que tenga por objetivo uno de los elementos de la lista estará prohibido.

Este precepto, al igual que el resto del capítulo sobre dirección algorítmica, se aleja del concepto restrictivo del art. 22 RGPD que exige para su aplicación que la decisión sea totalmente automatizada sin intervención humana[13]. Por el contrario, el art. 7.3 de la DTPD señala expresamente que la prohibición se aplica a las plataformas cuando usen sistemas automatizados para dar soporte o tomar decisiones que afectan a las personas de cualquier forma. De esta forma, aunque exista intervención humana significativa en el procesamiento o toma de decisiones relacionados con la lista de elementos prohibidos, la prohibición de uso de sistemas automatizados se mantiene.

[13] Todolí Signes, A., "La gobernanza colectiva de la protección de datos en las relaciones laborales: big data, creación de perfiles, decisiones empresariales automatizadas y los derechos colectivos", RDS, 84, 2018.

Respecto al ámbito subjetivo de la prohibición, el art. 7.2 establece que incluye procesar los datos, no solo de todas las personas prestando servicios en las plataformas (por tanto, con independencia de que sean asalariados o autónomos), sino también de aquellos potenciales prestadores de servicios que se encuentren en proceso de reclutamiento o de selección.

### *3.2. Trasparencia. Derechos de información y consulta*

En los últimos años, el principio de transparencia ha adquirido una relevancia central y omnipresente en todas las regulaciones tecnológicas. Específicamente, un estudio indica que el 94% de los documentos analizados sobre inteligencia artificial destacan la transparencia como un principio fundamental en cualquier normativa o desarrollo ético relacionado con la IA[14]. De manera similar, otro estudio muestra que, de 84 documentos sobre valores éticos en materia algorítmica, la transparencia es el principio más mencionado[15].

Estos principios éticos se han incorporado ampliamente en la normativa reciente. Por ejemplo, el Reglamento General de Protección de Datos (RGPD), el Reglamento Europeo de Inteligencia Artificial (RIA) incluyen el principio de transparencia, además de otros derechos que concretan este principio multifacético[16]. En el mismo sentido, la DTPD establece una completa regulación que tiene por objetivo reducir la asimetría informativa entre la plataforma digital y los prestadores de servicios –tanto asalariados como autónomos–.

A su vez, este grupo de preceptos otorgan derechos de información y consulta a los representantes de los trabajadores respecto al uso de estas tecnologías en la plataforma digital. El carácter colectivo de estos derechos de información y consulta sigue la estela pionera de la Ley Rider española[17] y es una novedad a nivel europeo. En efecto, la doctrina ha sido crítica con el RGPD por establecer solamente derechos de información individuales e incluso por hacer recaer el peso de la legitimidad del tratamiento de datos automatizado en el consentimiento individual en una relación marcada por una gran desigualdad entre las partes que hace difícil considerar el consentimiento como válido como es en el contrato de trabajo. Por ello, la doctrina ha apostado por "negociar el algoritmo"[18] a través de los

[14] Fjeld, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI", Berkman Klein Center for Internet & Society Research at Harvard University. 2020

[15] Jobin, A et al "The Global Landscape of AI Ethics Guidelines", *Nature Machine Intelligence*, 1, 2019, 389-399.

[16] Ver Todolí Signes, A., "El principio de transparencia algorítmica en su dimensión individual y colectiva: Especial referencia a la Directiva de plataformas digitales y al Reglamento de IA", Trabajo y Derecho, 19, 2024.

[17] Todolí Signes., A., "Cambios normativos en la Digitalización del Trabajo: Comentario a la "Ley Rider" y los derechos de información sobre los algoritmo", IUSLabor. Revista d'anàlisi de Dret del Treball, [en línea], 2021, n.º 2, pp. 28-65.

[18] A respecto ver, De Stefano, V.: "Negotiating the algorithm: Automation, artificial intelligence, and labor protection", *Comparative Labour Law and Policy Journal*, 41 (1), 2020, pp. 15-46; DE STEFANO, V. y

representantes legales de los trabajadores. La redacción final de la DTPD no alcanza a establecer obligaciones de negociación en el sentido estricto o de cogestión, pero sí exige que la implementación de estos sistemas automatizados –de control y de toma de decisiones– sean consultados con los representantes legales de los trabajadores.

La DTPD cuando concede derechos de información y consulta diferencia, en ocasiones, de un lado, los conceptos de representantes de los trabajadores y, de otro, representantes de las personas prestando servicios en plataformas. La diferencia viene provocada porque los derechos de información de carácter colectivo se conceden, no solamente en plataformas donde hay asalariados –y, por tanto, representantes legales clásicos–, sino también en aquellas plataformas donde los prestadores de servicios son verdaderos autónomos. La Directiva no establece ni regula sistemas de elección de representantes de los autónomos ni atribuye dichas funciones a órganos específicos creados por la DTPD, sino que concede el estatus de "representantes de las personas prestadores de servicios de plataformas" conforme a las leyes del EM o la práctica habitual en tales estados (art. 2.1 (g) DTPD.

Adicionalmente, la normativa al definir representantes legales de los trabajadores (art. 2.1 (f)) no parece dar preferencia a un tipo de representación u a otra. Por el contrario, conforme el considerando 22, los representantes son personas reconocidas como tal por las leyes de los EM o la práctica de cada país. Los únicos requisitos que se exigen es que sean representantes designados o elegidos por los sindicatos o elegidos como representantes por elecciones libres en las empresas conforme a la normativa nacional. Esta amplitud en la que la DTPD define quién puede ser representantes a efectos de los derechos de información y consulta será relevante para la trasposición, dado que permitirá que, en aquellos casos en los que no existan representantes legales de los trabajadores en la propia plataforma[19], estos sean suplidos por los sindicatos más representativos u otros mecanismos de representación[20]. También parece que la norma de trasposición podría dar preferencia a unos representantes sobre otros cuando existan los dos modelos en una concreta plataforma, puesto que la DTPD remite a la normativa del EM para concretar estos aspectos[21].

La única regulación que establece la normativa respecto a las relaciones entre diferentes tipos de representantes es la subsidiariedad de los representantes de los proveedores de servicios. En este sentido, el art. 15 DTPD establece que estos representantes

TAES, S.: "Algorthmic Management and Collective Bargaining." Transfer: European Review of Labour and Research, vil. 29(1), 2023, pp. 21-36; SÁEZ LARA, C.: "Gestión algorítmica empresarial y tutela colectiva de los derechos laborales." Cuadernos de Relaciones Laborales, 40 (2), 2022, p.297 y ss.; COLLINS, P. y ATKINSON, J.: "Worker voice and algorithmic management in post-Brexit Britain", Transfer: European Review of Labour and Research, Vol. 29 (1), 2023, p. 4

[19] O representantes de las "personas prestadores de servicios de plataformas".

[20] Como ya ocurre en la negociación de los planes de igualdad.

[21] Lo único que es obligatorio, de acuerdo con el art. 14 de la DTPD, es que en caso de que no haya representantes de los trabajadores en la plataforma, los EM deben asegurar que la plataforma digital de trabajo informe directamente a los trabajadores implicados.

podrán ejercer los derechos otorgados por la DTPD[22] a los representantes legales de los trabajadores, pero solamente en beneficio del personal autónomo. Esto es, la normativa parece querer evitar que, a falta de representantes legales de los trabajadores, los representantes de los proveedores de servicios cubran ese espacio ejerciendo los derechos de información y consulta a favor de los asalariados. La precaución podría tener su origen en un temor de legislador a la posible "captación" de estos representantes no sindicales (y de los que no existe una fuerte tradición en Europa) por parte de la plataforma digital. En cualquier caso, cabe señalar que el art. 22 de la DTPD dictamina la obligación de implementar medidas de protección, en favor de los los representantes, frente a las represalias de la plataforma por ejercer sus funciones.

El contenido de los derechos de información es muy detallado. Por un lado, se regula una extensa lista de aspectos que deben ser notificados a las concretas personas prestadoras de servicios (con independencia de su clasificación jurídica) y a sus representantes (art. 9. 1)[23]. La DTPD fija que esta información deberá entregarse por escrito y la información deberá presentarse de forma trasparente, inteligible, en un formato fácilmente accesible utilizando lenguaje claro y sencillo (art. 9.2 DTPD). También se regula el momento de entrega de la información: el primer día de trabajo o antes de la introducción de cambios que afecten a las condiciones de trabajo y también en cualquier momento en que sea solicitada la información (art. 9.3 DTPD). Adicionalmente, deberá entregarse toda la información a las personas en proceso de selección (art. 9.5). Cabe entender, en este sentido, que aunque las plataformas no hacen un proceso de selección al uso, la obligación de información nacerá en el momento anterior a que soliciten formar parte de la plataforma. Esto es, dentro del proceso a seguir decidido por la plataforma para formar parte de ella (y poder prestar servicios a través de ella o para ella) toda la información del art. 9.1 DTPD deberá estar disponible y accesible.

[22] Esto es, los derechos otorgados bajo el Artículo 8.2, el Artículo 9.1 y 9.4, el Artículo 10.4 y el Artículo 11.2 de la DTPD.

[23] Esa información deberá referirse a: (a) todos los tipos de decisiones respaldadas o tomadas por sistemas de toma de decisiones automatizadas, incluso cuando dichos sistemas respalden o tomen decisiones que no afecten de manera significativa a las personas que realizan trabajo en plataformas; (b) en cuanto a los sistemas de monitoreo automatizado: (i) el hecho de que tales sistemas están en uso o en proceso de ser introducidos; (ii) las categorías de datos y acciones monitoreadas, supervisadas o evaluadas por dichos sistemas, incluida la evaluación por parte del destinatario del servicio; (iii) el objetivo del monitoreo y cómo el sistema pretende lograrlo; (iv) los destinatarios o categorías de destinatarios de los datos personales procesados por dichos sistemas y cualquier transmisión o transferencia de dichos datos personales, incluida dentro de un grupo de empresas; (c) en cuanto a los sistemas de toma de decisiones automatizadas: (i) el hecho de que tales sistemas están en uso o en proceso de ser introducidos; (ii) las categorías de decisiones que son tomadas o respaldadas por dichos sistemas; (iii) las categorías de datos y los parámetros principales que tales sistemas tienen en cuenta y la importancia relativa de esos parámetros principales en la toma de decisiones automatizada, incluida la manera en que los datos personales o el comportamiento de la persona que realiza trabajo en plataforma influyen en las decisiones; (iv) los motivos de las decisiones para restringir, suspender o terminar la cuenta de la persona que realiza trabajo en plataforma, para rechazar el pago por el trabajo realizado por ella, así como para decisiones sobre su estatus contractual o cualquier decisión de efecto equivalente o perjudicial.

Adicionalmente a esta información, en el ámbito colectivo, el art. 13 DTPD determina que los representantes de los trabajadores en plataformas digitales de trabajo tendrán derecho de información y consulta –en los términos de la Directiva 2002/14/EC– cuando la plataforma tome decisiones que probablemente lleven a la introducción (o cambios sustanciales en el uso) de sistemas automatizados de seguimiento o de toma de decisiones.

Igualmente, la DTPD configura obligaciones adicionales en los casos en los que la plataforma digital deba realizar una evaluación de impacto conforme al art. 35 del RGPD. Concretamente, el art. 8 de la DTPD determina que en el desarrollo de la evaluación se deba consultar a las personas prestando servicios en la plataforma y sus representantes. Además, las plataformas deberán entregar la evaluación de impacto a los representantes de los trabajadores[24]. A mi juicio, el mecanismo de cumplimiento de este precepto debería ser similar a cómo se desarrollan los planes de igualdad. Es decir, la empresa debería presentar una propuesta de diagnóstico –que sería la evaluación– y con los representantes se negocia tanto el diagnóstico como las medidas para reducir el impacto de la tecnología. En fin, la conjunción de los art. 8 y 13 de la DTPD parecen implicar una intervención fuerte de los representantes en cualquier implantación de un sistema de seguimiento automatizado o toma de decisiones algorítmico que va más allá de la mera entrega de información.

### *3.3. Auditoria algorítmica y de los sistemas de control*

Bajo el principio fundamental de que una persona debe estar en control del algoritmo –siguiendo el extendido lema del "human in command"[25]– la DTPD establece obligaciones de supervisión (art. 10) y revisión (art. 11) sobre el algoritmo que deben ser realizadas por humanos. Sin embargo, a pesar del título de los preceptos, su contenido no se centra tanto en que una persona humana deba realizar las actividades de supervisión y revisión como que las mismas deban hacerse correctamente. Es decir, las obligaciones impuestas tienen por objetivo garantizar la no vulneración de los derechos de las personas prestadoras de servicios y no la simple intervención humana en el proceso. De ahí que, a mi juicio, sería más acertado que, al menos el art. 10 DTPD, se hubiera titulado "obligaciones de auditoría algorítmica".

Se busca conseguir este objetivo mediante el establecimiento de obligaciones *expost*, esto es, supervisión y revisión del resultado obtenido tras la toma de decisiones en la que interviene el algoritmo o el sistema automatizado de control. Así, la normativa dictamina obligaciones de reducción del riesgo de vulneración de derechos durante el desarrollo del sistema algorítmico y antes de su implantación en la empresa (el art. 35 RGPD y 8 y 13

[24] Este es uno de los supuestos en los que la Directiva da preferencia a los representantes de los trabajadores. Ahora bien, ante la ausencia de estos, el art. 15 de la DTPD, permite que la entrega de la evaluación de impacto se realice a los representantes de las personas proveedoras de servicios aunque solamente a efectos de sus representados que son las no asalariados.

[25] Un lema muy extendido entre la doctrina más autorizada y que fue incorporado al informe de la OIT sobre el Futuro del trabajo, Global Commission on the Future of Work, Work for a brighter future, ILO, 2019.

DTPD). Adicionalmente se establece la obligación de supervisar que el sistema, una vez implantado, funciona correctamente y no vulnera dichos derechos (art. 10 DTPD).

Concretamente, el art. 10 DTPD indica que las plataformas digitales deberán supervisar regularmente –máximo cada dos años– y realizar una evaluación del impacto de las decisiones individuales tomadas o realizadas con el apoyo de los sistemas automatizados de control o de toma de decisiones en sus condiciones de trabajo y en materia de igualdad. Así pues, como se acaba de defender, la norma no impone solamente la intervención humana en la supervisión de los efectos que esté provocando el sistema automatizado, sino que ordena realizar una evaluación del impacto a posteriori: lo que yo llamaría una auditoría algorítmica[26].

En su art. 10.2 DTPD se indica que los EM deben requerir a la plataforma digital para que tenga suficientes recursos para una efectiva supervisión y evaluación (auditoria). Además, se especifica que el responsable de la auditoria deberá tener suficiente competencia, formación y autoridad para realizar sus funciones incluida la autoridad competencial para modificar decisiones automatizadas individuales. Esta persona deberá tener protección frente a represalias por parte de la empresa por ejercer sus funciones.

Una cuestión relevante que surge es si la plataforma digital podría externalizar la realización de la auditoria en terceras empresas. Actualmente se está creando un mercado relevante de empresas especializadas en la realización de auditorías algorítmicas y tendría sentido que la empresa pudiera externalizar alguna de las partes de esa auditoria de impacto. Ahora bien, la literalidad del precepto implica que exista un responsable en la propia empresa al estilo del Delegado de Protección de datos –de hecho nada impide que fuera la misma persona–. La intención de que dicha persona o personas (la Directiva habla en plural) sea interna se vislumbra precisamente por concederle protección frente al despido (u otras represalias) por el ejercicio de sus funciones. A pesar de esto, considero que sería perfectamente posible que la empresa externalizara parte del proceso de auditoría bajo la supervisión o control de estas personas expertas y responsables nombradas por la plataforma digital.

Adicionalmente, la norma exige que dicha supervisión (auditoría) se realice con la participación de los representantes de los trabajadores (art. 10.1 DTPD). La norma no aclara cuál debe ser el nivel de intervención. Se debe entender que no puede ser un simple derecho de información y consulta sobre la auditoría, puesto que entonces el legislador hubiera usado ese término –que sí usa en el art. 13 de la propia Directiva–. De esta forma, la participación de los representantes en la auditoría debe ser algo más. En mi opinión, la solución puede venir por lo que algunas empresas ya están realizando dentro de la UE[27]: los comités conjuntos. En efecto, el responsable final de la realización de la

[26] Sobre la necesidad de obligar a establecer auditorías algorítmicas como forma real de proteger los derechos de las personas trabajadoras frente a la dirección algorítmica del trabajo ver, Todolí Signes, A., Algoritmos productivos y extractivos, Aranzadi, 2023.

[27] Rego, K.: "Works councils and the digitalization of manufacturing: Opportunity or threat for their power position?", *Economic and Industrial Democracy*, vol. 43 (4), 2022; De Stefano, V. y Taes, S.: "Algorthmic Management and Collective Bargaining." *Transfer: European Review of Labour and Research*, vil. 29(1),

evaluación será la plataforma, sin embargo, el proceso y la toma de decisiones como: la elección, en su caso, de la empresa a la que subcontratar, el tipo de auditoría a realizar, la intensidad de la misma, las medidas a tomar tras los resultados de la auditoría, etc., deberían adoptarse mediante un comité paritario en el que los representantes de los trabajadores formaran parte con las mismas potestades que los representantes de la empresa.

Finalmente, el art. 10.5 nombra la intervención humana, en su sentido más clásico, al indicar que todas las decisiones que impliquen restringir, suspender o terminar una relación contractual con una persona prestadora de servicios deben ser tomadas por un humano. Por la redacción, no cabe duda, de que la intervención humana en este tipo de decisiones deberá ser significativa. La redacción no deja dudas de que la mera convalidación de la decisión automatizada por un humano no cumpliría la exigencia de que la decisión sea "tomada" por un ser humano.

De hecho, el debate que puede surgir respecto a este contundente precepto es el contrario. Es decir, dilucidar si el precepto permite que este tipo de decisiones se tomen con apoyo de un sistema automatizado o, por el contrario, si el precepto prohíbe también cualquier tipo de intervención automatizado en este tipo de decisiones. Esta última opción no parece el objetivo de la Directiva. En general, tal y como expresa el considerando 48 de la Directiva, este precepto es asimilable –con una redacción más clara para reducir posibles problemas interpretativos– al art. 22 de del RGPD. En este sentido, el art. 10.5 DTPD tendría por objetivo garantizar una intervención humana significativa en la toma de este tipo de decisiones. Como ha indicado la doctrina, esto implica que la plataforma, para demostrar que hay intervención humana significativa, debería acreditar que, de forma estadísticamente relevante, se aparta de la toma de decisiones del algoritmo o, alternativamente, que el ser humano realiza un proceso de toma de decisiones independiente y que posteriormente es comparado con el tomado por el sistema automatizado[28].

Un riesgo que puede tener esta interpretación, que asimila el objetivo del art. 10.5 DTPD con el contenido del art. 22 RGPD, es concluir que la intervención humana significativa es solamente aplicable para decisiones de restricción, suspensión o despido de prestadores de plataformas digitales. Sin embargo, aunque el ámbito del 10.5 DTPD se refiere únicamente a estas materias, el art. 22 RGPD hace referencia a cualquier decisión que "produzca efectos jurídicos" sobre la persona. De esta forma, parece relevante señalar que otro tipo de decisiones automatizadas que no sean del contenido descrito en el art. 10.5 DTPD le seguirá siendo aplicable el art. 22 RGPD.

### *3.4. Derecho de revisión*

El art. 11 DTPD mantiene el objetivo de una IA "bajo control humano". No obstante, su alcance es mucho más concreto. El precepto comienza estableciendo el "derecho a una

2023, pp. 21-36; Collins, P. y Atkinson, J.: "Worker voice and algorithmic management in post-Brexit Britain", *Transfer: European Review of Labour and Research*, Vol. 29 (1), 2023, p. 4.

[28] Todolí Signes A., *Algoritmos productivos y extractivos*, Aranzadi, 2023.

explicación". Este derecho ya forma parte del RGPD, por lo que de nuevo es un precepto que tiene por objetivo aclarar y concretar la aplicación de este en plataformas digitales. En efecto, el "derecho a una explicación" contenido en el RGPD se encuentra sometido a un fuerte debate doctrinal entre posiciones muy alejadas. Por un lado, se defiende que la persona tiene derecho a una explicación previa de cómo funciona, en general, el sistema algorítmico[29]. Por otro lado, otra parte de la doctrina[30], entre las que me encuentro[31], sostienen que el derecho a la explicación nacido del RGPD frente a decisiones automatizadas es a posteriori. Esto es, se tiene derecho a conocer porqué el sistema algorítmico ha tomado la concreta decisión que ha afectado a la concreta persona que ha reclamado la explicación.

Este debate no se va a replicar respecto al art. 11 de la DTPD. Este artículo es claro al indicar que el derecho a obtener una explicación se refiere a la concreta decisión tomada por el sistema automatizado o con soporte del sistema automatizado. Adicionalmente, se indica que esa explicación debe presentarse sin retrasos indebidos, de forma trasparente, inteligible y usando lenguaje claro y sencillo. Es decir, la norma aclara que no se busca una explicación del funcionamiento del algoritmo (o una explicación matemática de su funcionamiento) sino de los motivos que han llevado a la toma de esa decisión. Esto es, de los comportamientos, actitudes o aptitudes de la persona, de las valoraciones, evaluaciones o ponderaciones realizadas por el algoritmo y del nexo causal que ha llevado al resultado concreto (la decisión algorítmica o el resultado usado para apoyar la decisión humana).

Cuando la decisión tomada es de restricción, suspensión o terminación de la cuenta de la persona prestando servicios, la normativa se vuelve más garantista. En este supuesto, la norma no configura un derecho subjetivo a una explicación que debe ser activado por la persona afectada –o sus representantes–, sino que instaura una obligación a la plataforma de entregar por escrito las razones de la decisión. Ahora bien, estas garantías son diluidas posteriormente, cuando el art. 11.4 DTPD determina que el contenido del art. 11 no afecta a los procedimientos disciplinarios o de despido establecidos por la norma nacional o los convenios colectivos. De esta forma, parece indicar que la obligación de entregar por escrito las razones de la restricción, suspensión o terminación de la

[29] Wacher, B., Mittelstadt Y Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Proteccion Regulation" International Data Privacy Law, 7, 2017, pp 79-90.

[30] Isak Mendoza & Lee A. Bygrave, "The Right Not to Be Subject to Automated Decisions Based on Profiling", en Tatiani Synodinou et al. (eds.,) EU internet law: regulation and enforcement, 2017, https://papers.ssrn.com/abstract=2964855 [https://perma.cc/XV3T-G98W]; Edwards L Y Veale M., "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For", Duke Law & Technology Review, 16, 2017, p 18-82; Wan Kim Y Routledge B., "Algorithmic Transparency, a Right to Explanation, and Placing Trust", Squarespace, 2017, https://static1.squarespace.com/static/592ee286d482e908d35b8494/t/59552415579fb3 0c014cd06c/1498752022120/Algorithmic+transparency%2C+a+right+to+explanation+ and+trust+%28TWK%26BR%29.pdf [https://perma.cc/K53W-GVN2, visitado 07/05/2024; Gianclaudio Malgieri & Giovanni Comandé, "Why a Right to Legibility of Automated Decision Making Exists in the General Data Protection Regulation", International data privacy Law, 7 (4) 243, 246-47, 2017.

[31] Todolí Signes, A., "La gobernanza colectiva de la protección de datos en las relaciones laborales: big data, creación de perfiles, decisiones empresariales automatizadas y los derechos colectivos", RDS, 84, 2018.

relación jurídica/la cuenta en la plataforma no deberá ser valorado –o al menos la Directiva no obliga a ello– a efectos de la calificación jurídica del despido o de otras sanciones disciplinarias. En cualquier caso, esto no parece que vaya a afectar a nuestro país donde la normativa ya exige que la carta de despido contenga las razones del mismo.

## 4. Recomendaciones para la trasposición

La Directiva otorga a los EM dos años para su trasposición y expresamente permite que en ella se puedan mejorar sus prescripciones en beneficio de las personas trabajadoras y autónomas (art. 26 DTPD). De esta forma, el Estado español tiene una oportunidad única para clarificar y desarrollar algunas de las cuestiones contenciosas que se plantean en la Directiva.

En primer lugar, considero que el contenido de esta Directiva debería ser traspuesto en el propio ET. En los últimos años se ha producido una disgregación normativa laboral al margen del Estatuto (ej., Ley del trabajo a distancia, Ley orgánica de protección de datos y garantía de Derechos digitales, entre otros). Como se ha sostenido en otros lugares[32], el fenómeno de la "huida" del Estatuto de los Trabajadores provoca problemas interpretativos, competenciales y de efectividad de la norma laboral. Estos problemas se verían reducidos si los derechos laborales se incorporaran directamente en el ET. Adicionalmente, en los últimos años a nivel político se ha planteado la necesidad "llevar" el ET al S.XXI. Difícilmente se entendería que se pretende cumplir este objetivo si la regulación de una parte esencial de lo que es el trabajo en el S.XXI, el trabajo en plataformas queda al margen de este[33].

En línea con lo anterior, la trasposición deberá aclara quién es el órgano competente para supervisar y controlar el cumplimiento de la normativa traspuesta. La DTPD, en su art. 24, señala que "Las autoridades de control responsables de supervisar la aplicación del Reglamento (UE) 2016/679 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 esta forma, en materia de protección de datos contenida en los art. del 7 al 11, el órgano competente deberá ser la AEPD. Sin embargo, este conjunto de artículos, del 7 al 11, contiene muchas prescripciones más allá de la mera protección de datos. Así, por ejemplo, cuando el art. 7.1 b) DTPD impide "tratar ningún dato personal relacionado con conversaciones privadas, tampoco los intercambios con otras personas que realicen trabajo en plataformas ni con los representantes de las personas que realizan trabajo en plataformas", lo que implica que el procesamiento de ese tipo de datos debería ser controlado y sancionado por la AEPD. Ahora bien, este tipo de procesamiento también implica una vulneración del derecho a la intimidad y en el caso de las conversaciones con los representantes legales determinaría una vulneración

[32] Todolí Signes, A., "Algoritmos productivos y extractivos", Aranzadi, 2023.

[33] Respecto a las garantías frente a los algoritmos concedidas a las personas naturales en plataformas digitales estas podrían incorporarse a la LOPDGDD o a la Ley del trabajo autónomo.

de la libertad sindical cuyo competente en materia de sanción debería ser la ITSS. En el mismo sentido, la existencia de una evaluación de impacto, art. 35 RGPD, debería ser controlada por la AEPD, pero el control de que la misma sea "consultada" con los representantes del personal y que sea entregada a estos, de nuevo, debería recaer en la ITSS. De esta forma, considero que la LISOS debería ser actualizada para concretar y aclarar la tipificación del incumplimiento de las salvaguardas frente a los algoritmos por razones de seguridad jurídica y garantía de la efectividad de la norma.

En segundo lugar, la normativa establece un amplio abanico de protecciones frente al control automatizado y a las decisiones automatizadas, unas garantías que, sin ser prefectas, mejoran sustancialmente las existentes en el RGPD, amén de clarificar muchos de los problemas interpretativos que el Reglamento genera en esta materia. En este sentido, considero que las protecciones frente a los algoritmos –art. 7 al 15 DTPD– deberían aplicarse a todas las empresas que usen sistemas de control automatizado o de toma de decisiones automatizadas y no solamente a las plataformas digitales. Cuatro razones justifican esta propuesta.

1. La normativa de protección algorítmica de la Directiva es técnicamente mejor, más completa y actualizada que la contenida en el RGPD. De un lado, el Reglamento fue aprobado en 2016. La rápida evolución de la IA en estos ocho años, así como un mejor conocimiento sobre los riesgos que plantea la IA para los derechos de las personas ha permitido una regulación actualizada. De otro lado, el RGPD es una norma trasversal pensada principalmente para proteger consumidores informados (de ahí que el consentimiento individual sea la principal forma de legitimación del tratamiento de datos), donde instituciones como la identificación de un responsable, la elaboración de normas técnicas, la previsión de mecanismos de control o la relevancia del consentimiento son sus señas identitarias. Por el contrario, los principios aplicables a las normas laborales son muy distintos: garantía de derechos, limitaciones a la autonomía de la voluntad, reconocimiento de la autonomía colectiva, legitimación del conflicto. En este contexto, la DTPD se acerca mucho más a la perspectiva laboral clásica, regulando específicamente para las necesidades del mercado de trabajo.
2. Los riesgos frente a la dirección algorítmica del trabajo analizados por la doctrina y que se pretenden reducir con las prescripciones contenidas en la DTPD no afectan solamente a los trabajadores de plataformas. Por el contrario, el hecho diferencial que hace necesarias las protecciones específicas es la existencia de un control automatizado o un sistema de toma de decisiones automatizado. Es decir, no es el factor "plataforma" lo que justifica la necesidad de estas garantías, sino el hecho de que un algoritmo tome decisiones que afecten a personas. Por esta razón, no hay justificación suficiente para excluir de la aplicación de estos derechos las personas trabajadoras sometidas a un algoritmo, pero que no trabajan en plataformas.

3. El tercer motivo, muy relacionado con lo anterior, es evitar la creación de trabajadores de primera y de segunda. Las diferenciaciones injustificadas en materia de protecciones laborales violentan el principio de igualdad ante la ley. De esta forma, si el factor diferencial es el control o la toma de decisiones automatizada este debería ser el factor activado/justificante de las protecciones.
4. La última razón es por seguridad jurídica. En efecto, ya con el RGPD, gran parte de las prescripciones frente a algoritmos que contiene la DTPD están incluidas y, por tanto, son obligatorias para todas las empresas y no solo las plataformas. Esto es, la DTPD solamente aclara y detalla las mismas. Así pues, convendría aplicar dicha "aclaración" a todas las empresas y no solo a las plataformas por razones de seguridad jurídica.

Por este argumento, considero que la trasposición debería incorporar estos derechos, no solo en el ámbito de plataformas, sino también para otras empresas siempre que usen esos medios tecnológicos para monitorizar, tomar o fundamentar decisiones organizativas o empresariales que afecten a trabajadores.

En tercer lugar, la trasposición debe concretar respecto a quien se predican los derechos colectivos incluidos. De esta forma, cabría aclarar si se da preferencia a las secciones sindicales o al comité de empresa y en qué supuestos. En cualquier caso, parece conveniente indicar que en caso de que no existan representantes legales en la plataforma estos derechos son reconocidos a los sindicatos más representativos como se ha hecho con los planes de igualdad.

En fin, se debe hacer una valoración positiva de la norma aprobada en el seno de la Unión. No obstante, en este trabajo se plantea una serie de dudas jurídicas que deberían quedar resultas en la trasposición con objeto de reducir la inseguridad jurídica y de garantizar que se cumple el objetivo normativo de mejora de las condiciones laborales de los trabajadores en plataformas digitales y sometidos a la dirección algorítmica.

# Gestión de algoritmos. El caso del trabajo en plataformas

## Managing the algorithms. The case of platform work

Nastazja.Potocka-Sionek
*Post-doctoral researcher. University of Luxembourg*
ORCID ID: 0000-0001-9841-5068

doi: 10.20318/labos.2024.9031

*Resumen:* El artículo analiza el primer instrumento de la UE para regular la supervisión automatizada y la toma de decisiones automatizada en el contexto laboral, es decir, la Directiva sobre la mejora de las condiciones de trabajo en el trabajo en plataformas (Directiva sobre el trabajo en plataformas). Las disposiciones legales sobre la gestión algorítmica contenidas en este instrumento merecen un análisis detallado. No solo se las considera ampliamente como el conjunto de disposiciones más progresistas y mejor diseñadas de la Directiva, sino también como un banco de pruebas para una mayor regulación que aborde las prácticas de gestión algorítmica en los lugares de trabajo tradicionales, más allá del contexto del trabajo en plataformas. El artículo analiza las disposiciones pertinentes establecidas en la Carta III de la Directiva sobre el trabajo en plataformas, prestando especial atención a la intrincada forma de establecer su alcance personal y material. La regulación actual de la supervisión y la toma de decisiones automatizadas en esa Directiva se contextualiza con las disposiciones de la original presentadas por la Comisión Europea y otros instrumentos legales pertinentes, como el Reglamento General de Protección de Datos. El artículo plantea que, a pesar de los notables avances en la protección de las personas que realizan trabajos en plataformas frente a los riesgos algorítmicos, algunos aspectos críticos siguen sin abordar.

*Palabras clave:* Trabajo en plataformas, gestión algorítmica, sistemas automatizados, protección de datos, condiciones laborales

*Abstract:* The paper provides an analysis of the firstever EU instrument to regulate automated monitoring and automated decisionmaking in the work context, i.e., the Directive on Improving Working Conditions in Platform Work (the Platform Work Directive). The legal provisions on algorithmic management contained in this instrument merit detailed scrutiny. Not only are they widely considered to be the most progressive and well-designed set of the Directive's provisions but also a testbed for further regulation that would address algorithmic management practices in traditional workplaces, beyond the platform work context. The article analyses the relevant provisions laid down in Charter III of the Platform Work Directive, paying particular attention to the intricate way of drawing their personal and material scope. The current regulation of automated monitoring and decision-making in that Directive is contextualised against the provisions of the original Proposal put forth by the EU Commission, as well as

*nastazja.potocka-sionek@uni.lu

other relevant legal instruments, such as the General Data Protection Regulation. The article posits that despite noticeable advancements in the protection of people performing platform work against algorithmic risks, some critical aspects remain unaddressed.

*Keywords:* Platform work, algorithmic management, automated systems, data protection, working conditions.

## 1. Introduction

Algorithmic management is a core driver of the platform business model and a definitional feature of digital labour platforms. As stems from the criteria formulated in Article 2 (1) Platform Work Directive (PWD), a digital platform does not fall under the scope of this instrument unless automated monitoring or automated decision-making systems are put in place.[1] Automated decision-making and monitoring shape the dynamics of platform work and are key determinants of the working conditions of people performing platform work. Relatedly, the regulation of algorithmic management lies at the very heart of the Directive, along with the presumption of the employment status of platform workers that tackles the perennial issue of their misclassification (Articles 4-6).

The algorithmic management provisions contained in Chapter III of the Directive (Articles 7 to 15) serve multiple purposes. From workers' perspective, the improvement of the protection of their personal data, regardless of their employment status, should shield them from excessive surveillance and enable them to realise their substantive rights, such as the right to privacy, the right to healthy and safe work conditions, and equal treatment.[2] The opacity of the algorithms used by the platform, and the information asymmetries between the platforms and their users, are key factors hindering the full exercise of these labour rights. Moreover, information and data access rights should make workers aware of the mechanisms steering their work performance, which is instrumental to regaining control over the working process, and making it more predictable and easier to navigate. Data rights are also pivotal to identifying potential biases and claiming labour rights before courts and national authorities (i.e., data protection authorities, equality bodies and labour authorities). More broadly, increasing transparency of algorithmic systems should also improve legal certainty and ensure a level playing field between digital labour platforms and offline providers. This is strongly related to the question of employment status, since the exercise of managerial functions and control through automated decision-making and monitoring has been recognised as one of the indicative criteria of a subordinate employment relationship. Thus, the regulation of algorithmic management has much broader implications than 'solely' protecting the personal data of people providing their

[1] Article 2 (1) (a) Platform Work Directive defines a digital labour platform as a 'natural or legal person providing a service which meets all of the following requirements: (i) it is provided, at least in part, at a distance through electronic means, such as a website or a mobile application; (ii) it is provided at the request of a recipient of the service; (iii) it involves, as a necessary and essential component, the organisation of work performed by individuals in return for payment, irrespective of whether that work is performed online or in a certain location; (iv) it involves the use of automated monitoring or decision-making systems'.

[2] Recital 4 PWD.

services through digital labour platforms. Rather, it is strictly related with the other goal of the Directive, namely the improvement of working conditions of platform workers. Even if the regulatory intervention in these fields has a different legal basis (Article 16 TFEU and Article 153(1) (b) TFEU respectively), these objectives mutually reinforce each other, and one is not subordinate to the other (Recital 16).

The term algorithmic management is not a statutory but primarily a doctrinal one. It is often attributed to Lee and colleagues, who were first to refer to it in 2015, describing it as a practice where the algorithms perform functions normally executed by human managers.[3] Almost ten years into the heated, interdisciplinary debate on this sociotechnical phenomenon, there is no commonly agreed and comprehensive definition of algorithmic management in literature. Some taxonomies focus more on the organisational coordination of labour,[4] others on monitoring and control,[5] and yet others on the whole spectrum of managerial prerogatives.[6] Still, the juxtaposition of algorithmic versus human management remains at the core of most conceptualisations[7] and is echoed in the Preamble of the Directive, which speaks of 'algorithms increasingly replace[ing] functions that managers usually perform in businesses.'[8] The 'techno-human dualism'[8] underlying this definition does not imply, however, a techno-deterministic or techno-centric view on algorithmic management, whereby the role of a human is secondary to technology. To the contrary, the Directive proposes a set of human-in-the-loop safeguards, which corresponds to the 'techno-human entanglement' approach premised on the assumption that humans can effectively mediate technological impact on the labour process.[9]

---

[3] LEE, Min Kyung, KUSBIT, Daniel, METSKY, Evan, and DABBISH, Laura. Working with machines: The impact of algorithmic, data-driven management on human workers. Proceedings of the 33rd Annual ACM SIGCHI Conference, Seoul, South Korea, 2015.

[4] BAIOCCO Sara, FERNANDEZ-MACÍAS Enrique, RANI Uma, PESOLE Annarosa. The Algorithmic Management of work and its implications in different contexts. *Background Paper Series of the Joint EUILO Project "Building Partnerships on the Future of Work"*. Geneva, 2022.

[5] E.g., Duggan and colleagues (2019) define it as a system of control whereby algorithms are responsibile for taking decisions affecting workers, diminishing human involvement and oversight of the labour process. See DUGGAN, James, SHERMAN Ultan, CARBERY Ronan, and MCDONNELL, Anthony. Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. Human Resource Management Journal, 30(1), 114-132.

[6] Wood conceptualised AM as 'software to automate organisational functions traditionally carried out by human managers'. See WOOD, Alex J. Algorithmic management consequences for work organisation and working conditions, JRC Working Papers Series on Labour, Education and Technology 2021/07, European Commission, 2021.

[7] Mateescu and Nguyen understand algorithmic management as a set of tools and technological processes that manage workers through digital means, replacing humans who direct and supervise workers with technology. See MATEESCU, Alexandra and NGUYEN, Aiha. Explainer: Algorithmic management in the workplace. 2019. Available at https://datasociety.net/wpcontent/uploads/2019/02/DS_Algorithmic _Management_E xplainer.pdf Recital 8 PWD.

[8] SULLIVAN Rick, VEEN Alex, and RIEMER Kai. Furthering engaged algorithmic management research: Surfacing foundational positions through a hermeneutic literature analysis. Information and Organization Volume 34, Issue 4, December 2024, 100528.

[9] Ibid.

The main goal of this paper is to assess in how far the measures laid down in the Platform Work Directive, adopted by the Council on the 14th of October 2024,[10] can be successful in shaping of a fair and equitable power relationship between platforms and people performing work through them, ultimately improving their working conditions.

## 2. An analysis of the 'algorithmic management' provisions of the Platform Work Directive

### *2.1. Limitations on processing of personal data*

Chapter III of the Platform Work Directive is opened by Article 7, which draws important red lines by limiting the material and temporal scope of collecting and processing personal data and delineating its purposes. First, it puts a ban on the processing of any personal data on the emotional or psychological state (Article 7 (1) (a)), and data in relation to private conversations (Article 7 (1) (b)). Second, it prohibits data collection while the person performing platform work is not working or offering to work (Article 7 (1) (c)). Third, it does not allow data processing to make predictions about the exercise of fundamental rights (Article 7 (1) (d)). Neither does it let it be applied to infer certain protected characteristics of workers, such as racial or ethnic origin, migration status, political opinions, religious or philosophical beliefs, disability, state of health, including chronic disease or HIV status, the emotional or psychological state, trade union membership, a person's sex life or sexual orientation (Article 7 (1) e)). Finally, it excludes the processing of biometric data to establish that person's identity by comparing that data to stored biometric data of individuals in a database (Article 7 (1) (f)). For the definition of biometric data, the Platform Work Directive refers to the General Data Protection Regulation, which defines it in Article 4 point 14 as 'personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data'.[11] Overall, Article 7 PWD has considerably expanded the catalogue of prohibited grounds of processing of personal data as compared to the original Proposal made by the EU Commission in December 2021. It has substituted a more general formula in that Proposal, prohibiting platforms from processing any personal data concerning platform workers 'that are not intrinsically connected to and strictly necessary for the performance of the contract between the platform worker and the digital labour platform'.[12] The final text built upon the GDPR and fundamental rights instrument (e.g., Article 8 of the Charter of Fundamental Rights

[10] The present analysis is based on the text available at https://data.consilium.europa.eu/doc/document/PE-89-2024-INIT/en/pdf.

[11] See also the definition of biometric data in the Recitals 42-43 PWD.

[12] Article 6 (5) of the Commission Proposal.

of the EU), contributing to the legal certainty and consistency of the legal framework. Making these requirements more specific renders their operationalisation easier.[13]

In addition to the red lines included in Article 7 PWD, Recital 39 states that digital labour platforms should not process the personal data of persons performing platform work on the basis that they have given consent to it. This is motivated by the fact that persons performing platform work do not have a genuinely free choice and cannot refuse or withdraw consent without detriment to their contractual relationship, because of the imbalance of power embedded in the platform-mediated work relationship. This corresponds with the provision of the Preamble to the GDPR, and with the interpretation of Article 29 WP Opinion, according to which the power asymmetry between the employer and the employee makes it highly unlikely that an employee's consent is freely given, without fear or experience of negative repercussions of a refusal of such consent.[14] As concluded by Article 29 WP, consent cannot be the legal basis in the majority of cases.[15]

Notably, Article 7 of the Platform Work Directive has a remarkably broad scope, as it applies not only to automated decision-making and automated monitoring as defined in Article 2 of this instrument but also to 'automated systems supporting or taking decisions that affect persons performing platform work in any manner'. This implies that also systems that do not affect them 'significantly', but even to a minor extent, are covered (Article 7 (3)).

Moreover, the personal scope of this provision is exceptionally broad. Not only is it detached from the employment status, covering people performing platform work without an employment contract, but it also encompasses those who undergo a selection procedure (Article 7 (2)). This extension is crucial given the impact of predictive HRM algorithms on the selection process, which is increasingly carried out without human intervention (e.g., automated admission to the online platform or task assignment based on selected worker characteristics).[16]

The next provision of the Directive, Article 8, has not been originally included in the Commission's Proposal. This new addition specifies the rule of Article 35 GDPR, which provides that processing of personal data by a digital labour platform by algorithmic management likely results in a high risk to the rights and freedoms of natural persons within the meaning of Article 35 GDPR. Digital labour platforms are therefore

[13] GUGLIELMETTI, Mario. Automated work and workers' rights: platform work and AI work management systems. In Artificial intelligence, labour and society, Aída Ponce Del Castillo (ed.), European Trade Union Institute (ETUI), Brussels, 2024. Available at https://www.etui.org/sites/default/files/2024-03/Artificial%20intelligence%2C%20labour%20and%20society_2024.pdf#page=129, p. 130.

[14] Opinion 10/2006 on the processing of personal data by the Society for Worldwide Interbank Financial Telecommunication (SWIFT), available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2006/wp128_en.pdf

[15] See also the Preamble of the GDPR, which states that a clear imbalance between the data subject and the controller means that consent should not be considered a valid legal basis for the processing of data.

[16] DUGGAN, James, SHERMAN, Ultan, CARBERY, Ronan, MCDONNELL, Anthony. Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. Human Resource Management Journal, 30(1), 2020, 114-132.

obliged to carry out a data protection impact assessment (DPIA) in line with Article 35 GDPR. In brief, a DPIA is a process designed to identify risks arising out of the processing of personal data and to minimise these risks as far and as early as possible. It is also specified that, in doing so, platforms shall seek the views of persons performing platform work and their representatives and that the results of DPIA shall be made available to workers' representatives. The explicit mention of the collective dimension is a significant step forward and has long been advocated for by scholars.[17] Details of the consultation and information procedure in this regard have not been provided, however.

### *2.2. Transparency in algorithmic management systems*

Next, Article 9 of the Directive moves on to provisions mandating transparency in algorithmic management systems. This provision lays down the fundaments for realising the goals of the Directive, since transparency is known as 'the first step towards genuine accountability.'[18] The personal scope of this Article is broad, covering persons performing platform work (regardless of employment status) and persons undergoing the recruitment and selection procedure as regards the automated monitoring and decision-making systems applicable to them (Article 9 (5)). Transparency rights also have a collective dimension, as they are granted to workers' representatives, who shall receive comprehensive and detailed information about all relevant systems and their features. Such information should be shared with them prior to the use of those systems or to the introduction of changes affecting working conditions, the organisation of work or monitoring work performance and, 'at any time upon their request'. Competent national authorities shall likewise obtain comprehensive and detailed information at any time upon their request.

The material scope of this provision is very comprehensive. Digital labour platforms are obliged to disclose information about the use of automated monitoring and automated decision-making systems as regards all types of decisions supported or taken by ADMSs, even when the algorithmically driven decisions do not affect persons performing platform work in a significant manner (9 (1) (a)). Article 9 (1) provision fills an important gap of the GDPR, under which data controllers are not obliged to inform subjects about the existence of algorithms which, according to them, are not covered by Article 22(1) GDPR, as long they comply with the information obligations under Articles 13-15 GDPR.[19] Only if the decision is fully automated and produces legal

[17] E.g., ADAMS-PRASSL, Jeremias. The Challenges of Management by Algorithm: Exploring Individual and Collective Aspects. Gyulavári, Tamás , and Emanuele Menegatti, ed. Decent Work in the Digital Age: European and Comparative Perspectives. Oxford: Hart Publishing, 2022. Bloomsbury Collections, p. 241.

[18] PONCE DEL CASTILLO, Aída and NARANJO, Diego. *Regulating algorithmic management An assessment of the EC's draft Directive on improving working conditions in platform work.* ETUI Policy Brief 2022.08, available at https://bit.ly/4f5y9sj

[19] HIESSL, Christina. Case Law on Algorithmic Management at the Workplace: Cross-European Comparative Analysis and Tentative Conclusions. 2023, p. 38. Available at https://ssrn.com/abstract=3982735

effects or similarly significant effects within the meaning of Article 22(1) GDPR will the information duties under that instrument be triggered. The Platform Work Directive, in turn, does not set either of these requirements, making it possible for platform workers' to access information on algorithmic management systems not contingent on the level of their automation. This broad scope overcomes the need for nuanced, technical debates on whether automation of AMS and ADMS on a given platform is full, partial, or conditional (Article 9 (1)).[20] Under Article 9 PWD, besides the information about the very fact that such systems are in use or are being introduced; the information shall cover the categories of automated or semi-automated decisions, categories of data and main parameters that such systems take into account, as well as the 'relative importance' of such parameters. This should include an information about how the personal data and behaviour of the person performing platform work influence the (semi-)automated decisions. Moreover, the grounds for certain decisions, i.e., about the account restriction, suspension or termination, the payment refusal after the work performance, and about their contractual status or equivalent should be made available. As regards automated monitoring systems, persons performing platform work should be informed about their usage and/or introduction; the categories of data and actions subject to this kind of monitoring, including evaluation by the end user; the aim of the monitoring and how it should be achieved; and the (categories of) recipients of the personal data processed by such systems, and its potential transmission or transfer. All this information about automated decision-making and automated monitoring systems should be provided in a written form and be easy to access and comprehend (Article 9 (2)). Moreover, at the latest on the first working day, before the introduction of changes, or upon their request, platforms shall inform people performing work through them about the systems and their features that directly affect them and their working conditions. The information shall be 'concise' but, upon their request, 'comprehensive and detailed' (Article 9 (3)).

Finally, Article 9 (6) grants persons performing platform work the right to the portability of personal data, including ratings and reviews. Digital labour platforms shall give them tools to facilitate the effective exercise of their portability rights, referred to in Article 20 of the General Data Protection Regulation. If a person performing platform work requests it, platforms shall transmit such personal data directly to a third party. This right is yet another valuable addition to the text of the Commission's proposal, which was not included in the original text, as it was deemed to be disproportionately burdensome for platform businesses, especially for small companies.[21] Ensuring intero-

[20] On the level of automation of algorithmic management in digital platform work, see e.g., BAIOCCO, Sara, FERNANDEZ-MACÍAS Enrique, RANI, Uma, and PESOLE Annarosa. The Algorithmic Management of work and its implications in different contexts. *Background Paper Series of the Joint EU-ILO Project "Building Partnerships on the Future of Work"*. Geneva, 2022, p. 6. As observed by the authors, 'full automation' would be technically possible only in the case of general AI, and even the most advanced models of algorithmic management applied by digital labour platforms require extensive human intervention at various stages.

[21] COMMISSION STAFF WORKING DOCUMENT IMPACT ASSESSMENT REPORT Accom-

perability of platforms' ratings should counter the so-called 'lock-in effect', which creates a dependency of platform workers on a given platform, preventing them from pursuing a career on multiple digital labour platforms independently.

Overall, Article 9 provides detailed and comprehensive safeguards regarding the so-called *ex-ante* transparency of algorithmic management systems.[22] The best illustration of the pertinence of workers' information and data access rights are cases decided by data protection authorities and courts in this regard, based on the GDPR. An example is the case of drivers working through the Ola platform, ruled by the Amsterdam Court of Appeal on the 4th of April 2023.[23] Drivers, supported by the App Drivers & Couriers Union and Worker Info Exchange, demanded access to their personal data processed by the platform, e.g., customer transactions, booking cancellation history, booking acceptance history, ratings, and GPS data for each trip. Their claim was based on Article 15(1) (h) GDPR, which provides for the right to be informed of the existence of automated decision-making, including profiling; the right to meaningful information about the logic involved; and the right to be informed of the significance and the envisaged consequences of such processing for the data subject. In this case, this article proved to be a powerful tool granting them access to multiple categories of data, such as 'device data' (i.e., data on the mobile phones used to carry out their rides), 'fraud probability score', or their 'earning profile'.

Still, this and similar cases demonstrate significant limitations of the GDPR in litigating workers' data rights.[24] Firstly, as mentioned above, workers can exercise their rights to data access in algorithmic management under Article 15 (1) (h) GDPR only if the decision is fully automated and produces legal effects or similarly significant effects within the meaning of Article 22(1) GDPR. If an algorithmic decision is made with human involvement (which is most typically the case), or a fully automated decision does not have a legal or significant effect, the algorithmic transparency requirement under Article 15(1)(h) does not apply. These criteria may give rise to interpretative doubts, as was the case in the Ola judgment.[25] Secondly, the extent of algorithmic management

panying the document Proposal for a Directive of the European Parliament and of the Council On improving working conditions in platform work. Available at https://data.consilium.europa.eu/doc/document/ST-14450-2021-ADD-2/en/pdf.

[22] VEALE, Michael, SILBERMAN, Michael 'Six' and BINNS, Reuben, Fortifying the algorithmic management provisions in the proposed Platform Work Directive. *European Labour Law Journal*, 14(2), 2023, p. 331.

[23] Case Number 200.295.806/01. An unofficial English translation of the judgment is available at https://5b88ae42-7f11-4060-85ff-4724bbfed648.usrfiles.com/ugd/5b88ae_de414334d89844bea61deaaebedfbbfe.pdf.

[24] For a comprehensive overview of cases concerning algorithmic management at work, see HIESSL, Christina. Algorithmic Management in the Workplace: Taking Stock of Case Law and Litigation in Europe. Hungarian Labour Law e-Journal. 2022/2, available at https://hllj.hu/letolt/2022_2_a/01_ChHiessl_hllj_uj_2022_2.pdf.

[25] For example, the 'fraud probability score' was qualified in the 1st instance as a District court to be 'profiling' within the meaning of Art. 4 (4) GDPR, but the drivers had not shown that automated decisions had been taken on the basis of this risk profile. The Court of Appeals reversed that decision and held that only the Guardian system, which was used to detect irregularities, did not qualify as a solely automated system, after Ola provided sufficient evidence that passengers would be contacted by staff.

information remains unclear, in particular concerning the interpretation of what constitutes 'meaningful information about the logic involved and the significance and the envisaged consequence'. Thirdly, workers' right to access in the context of algorithmic management can be limited by the intellectual property exception (Recital 63 of the GDPR) and by employers' protection against 'manisfestly excessive' requests (Article 12(5) GDPR). A case in point is the case decided by the District Court of Amsterdam, in which Uber demanded the applicants to specify the personal data they wanted to receive. The platform's request was approved by the Court, which found workers' request for the right of access request to be excessively general and not sufficiently specified.[26] As pointed out by Abraha, the fact that employers can ask workers to specify the data they demand can significantly limit workers' right of access in the context of algorithmic management, since usually they have a limited understanding of the collected and processed categories of personal data.[27] The Platform Work Directive specifies the categories of data to which workers should be granted access to, thereby filling critical gaps under GDPR and increasing legal certainty with regard to platforms' transparency obligations.

### *2.3. Human oversight*

In Article 10, the Platform Work Directive shifts the focus from *ex-ante* transparency measures of automated systems to their human oversight. It provides that platforms shall oversee and evaluate the impact of individual decisions taken or supported by algorithmic management systems used by the digital labour platforms on persons performing platform work. This should consider the impact of automated systems on the working conditions of people working through platforms, as well as their right to equal treatment at work. Should a high risk of discrimination or infringement of rights by automated decision-making or automated monitoring systems be identified, the platform is obliged to put in place necessary steps to avoid such decisions in the future. This may include, 'if appropriate', a modification or discontinuation of the algorithmic management systems. The person in charge of that process shall be adequately qualified and trained and, enjoy all the authority including the power to override automated decisions, and be protected from disciplinary measures or adverse treatment for exercising their functions. The oversight and evaluation process should take place regularly, at a minimum of every two years. This threshold might not be sufficient given the high pace of tech development; it would have been more appropriate to mandate the evaluation of the impact of individual decisions taken or augmented by automated systems every time a new system is introduced.

Article 10 PWD is certainly a much-needed recognition of the potentially negative impact of algorithmic management on working conditions and, in particular, of the

[26] Amsterdam Court of Appeal, Case number 200295.747/01. Unofficial English translation is available at https://5b88ae42-7f11-4060-85ff-4724bbfed648.usrfiles.com/ugd/5b88ae_21d84f102fee4f3888efcec9c

[27] ABRAHA, Halefom. Regulating algorithmic employment decisions through data protection law. *European Labour Law Journal*, *14*(2), 2023, p. 179.

die-hard problem of discrimination in platform work. There was no explicit mention of the risk of discrimination in the original Proposal by the Commission, which makes this addition an important advancement towards countering inequalities in platform work. One of the well-documented risks in this context is algorithmic wage discrimination, described by Dubal as a wage pricing technique whereby individual workers are paid differently based on intransparent calculations driven by data on location, individual behavior, demand, and supply, among other factors. It not only encompasses remuneration for completed work but also all decisions on the allocation of work and working time.[28] Another often-studied example is gender inequality, which is known to be reproduced and institutionalised by platforms' design choices and affordances.[29] This is an important gap given that discrimination, e.g., on grounds of gender or age, often occurs precisely at the time of selection. Cases of discrimination in digital labour markets are also documented to occur at the hiring stage.[30] Regrettably, however, human oversight under Article 10 PWD does not extend to people undergoing the recruitment procedure, as in the case of red lines with data processing under Article 7, but is limited to detecting discrimination while work is performed.

Workers' representatives also have a role to play in the human oversight procedures: they shall be involved in the oversight and evaluation procedure (Article 10 (1)) and obtain information on the evaluation of algorithmic systems (Article 10 (4)). Moreover, platforms shall make this information available to persons performing platform work and the competent national authorities upon their request. An important limitation of the effectiveness of this provision is the lack of external auditors to control whether the adjustments taken by the platform in the case of rights are adequate. National authorities have not been given any special competencies in this regard- they can only request information on the outcome of the oversight and evaluation procedure.

Finally, Article 10 (5) PWD lays down a special regime for decisions on restriction, suspension or termination of the contractual relationship or the account of a person performing platform work, mandating that it shall be taken by a human being (Article 10 (5)). This is a much-needed response to calls from experts to introduce a clear ban on 'robo-firing',[31] and a step forward as compared to the GDPR, where the prohibition of fully automated decisions that produce legal or similarly significant legal effects may be limited by a 'contractual necessity' clause (Article 22(2)).[32] At the same time, it is

[28] DUBAL, Veena. On algorithmic wage discrimination. *Columbia Law Review*, 123(7), 1929-1992. 2023.

[29] RENAN BARZILAY, Arianne. The Technologies of Discrimination: How Platforms Cultivate Gender Inequality. *The Law & Ethics of Human Rights*, vol. 13, no. 2, 2019, 179-202.

[30] FIERS, Floor. Inequality and discrimination in the online labor market: A scoping review. *New Media & Society*, *25*(12), 2023, 3714-3734.

[31] PONCE DEL CASTILLO, Aída. Regulating algorithmic management in the Platform Work Directive: correcting risky deviations. Global Workplace Law & Policy. Available at https://global-workplace-lawand-policy.kluwerlawonline.com/2023/11/22/regulating-algorithmic-management-in-the-platform-workdirective-correcting-risky-deviations/

[32] RAINONE, Silvia and ALOISI, Antonio; The EU Platform Work Directive What's new, what's missing, what's next? ETUI Policy Brief, 2024.06, August, available at bit.ly/4eIMYRA.

regrettable that only these three categories of decisions have been elevated to that status, unlike other important decisions impacting the contractual status and working conditions, although it is surely one of their main vulnerabilities.

### *2.4. Human review*

Article 11 PWD moves on to provisions on human review, sometimes referred to as '*ex post* transparency'.[33] First, persons performing platform work have the right to be informed about any decision taken or supported by an automated decision-making system. That explanation shall be presented without undue delay, orally or in writing, in a transparent and intelligible manner. Moreover, anyone performing platform work should have access to a designated contact person who can clarify the factors that have led to the decision.

Further, some categories of algorithmic decisions are subject to a stricter transparency regime, which requires a written statement of reasons provided without undue delay, at the latest on the day when it takes effect. This concerns decisions on the account restriction, suspension or termination; payment refusal, contractual status, as well as 'any other decision affecting the essential aspects of the employment or other contractual relationships.' Unlike in the original Proposal of the Platform Work Directive, the catalogue of decisions on the essential aspects of the work relationship between platforms and people performing platform work is open. While the requirement of a written form for those decisions is an important safeguard, the protection could have been stronger by requiring a notice period for changes to essential aspects of the contract, in particular in the case of dismissal.[34]

Besides the information on the (semi-) automated decisions, persons performing platform work shall be able to request a review of such decisions, including decisions that do not concern essential aspects of their contract. The same right is granted to representatives acting on behalf of the persons performing platform work with regard to personal data (Article 11 (2) in connection with Article 15). In response, platforms shall formulate a written, 'sufficiently precise and adequately substantiated reply' reply and communicate it without undue delay, maximum within two weeks of the request's receipt. It is worth pointing out that the two-week is an extension of the previously one-week period, which could have been extended only to micro, small, or medium-sized enterprises.

This human review is an important tool for workers and their representatives, in case they are dissatisfied with the explanation or the written statement of reasons ob-

[33] VEALE, Michael, SILBERMAN, Michael 'Six' and BINNS, Reuben. Fortifying the algorithmic management provisions in the proposed Platform Work Directive. *European Labour Law Journal*, 14(2), 2023.

[34] For example, Ontario's Digital Workers' Rights Act 2022 provides that 'where a platform worker's account is restricted, suspended or terminated by an automated decision-making system, the grounds for such a decision must also be made available to the platform worker, and the worker must be provided with two weeks' written notice of removal prior to a removal access. See OGUNDE, Fife. Algorithmic management of platform workers: An examination of the Canadian and European approaches to regulation. *European Labour Law Journal*, 2024, p. 11.

tained. The *ex-post* explanations of specific model outputs (the so-called 'local explanations') are complementary to the 'ex-ante', 'global' explanations concerning the general functioning of algorithmic mechanisms, enshrined in Article 9 PWD.[35] While both types of explanations are useful, qualitative research suggests that workers attach more importance to the 'local', concrete explanations and human review than to gaining access to general data. For example, a study of platform workers in the UK showed that while only 12 percent reported they would need access to info on platform AI or algorithmic usage and a right to request a personal and understandable explanation, human review of automated decision-making systems was indicated by 22 percent of respondents as a useful one.[36] Scholars, practitioners, and stakeholders have long voiced concerns about the risk that the information categories of data will not be fully understandable to those concerned, and have perceived the ex-post transparency obligations as more effective than the more abstract information rights concerning the systemic features of the algorithms applied by the platforms.

Article 11 (3) PWD mandates a digital labour platform to rectify any rights infringement caused by automated decision making or monitoring systems. Should a rectification not be possible, an adequate compensation should be offered. Platform's intervention may include a modification of the automated decision-making system or a discontinuance of its use. The reaction should be without delay and in any case within two weeks of the adoption of the decision. The obligations in this regard mirror those with regard to human overview set out in Article 10 (3) (see above). Article 11 does not specify which rights infringement is referred to. This provision should be interpreted broadly, as covering at least fundamental rights protected by the Charter of Fundamental Rights of the EU,[37] but not being limited to them.

The requirement to take necessary steps to avoid rights-infringing provisions has not been envisaged in the original Directive proposal put forward by the Commission.It constitutes a crucially important addition. What is missing, however, are provisions concerning reporting mechanisms and oversight over the measures taken by the platform to amend a deficient ADM or AMS, much as in the case of the detection of risk of discrimination or negative impact on working conditions under Article 10. The decision about the appropriate steps is entirely within the platforms' discretion. Regrettably, the proposal made in the European Parliament Report, which specified that in case an impact assessment found non-compliance with workers' rights and health and safety protection, data protection, labour and other competent authorities shall take coordinated measures

[35] VEALE, Michael, SILBERMAN, Michael 'Six', BINNS, Reuben, Fortifying the algorithmic management provisions in the proposed Platform Work Directive. *European Labour Law Journal*, 14(2), 2023.

[36] MARTINDALE, Nicholas, WOOD, Alex J., and BURCHELL, Brendan. What do platform workers in the UK gig economy want? *British Journal of Industrial Relations* 62(3), 2024, 542-567.

[37] The Directive refers to the several provisions of the Charter of Fundamental Rights of the EU in its Preamble (Recital 2), i.e., Article 31 on the right of every worker to fair and just working conditions; Article 27 on workers' right to information and consultation within the undertaking; Article 8 on the right to the protection of personal data; Article 12 on the right to freedom of assembly and of association, Article 16 on the freedom to conduct a business, and Article 21 on the right to non-discrimination.

to enforce those provisions, has not been followed. As expressed in that Report, concurrent supervision (i.e,. cooperation between the authorities as regards oversight) and the cumulative applicability of GDPR and labour law forms of redress, would be essential.[38] In the final version of the Directive, instead, the provisions on cooperation between data protection supervisory authorities and other competent authorities seem to focus only on information exchange. The provision would have been considerably strengthened by connecting it at least with a correlated obligation to declare the action taken to repair the system, submitted to the authorization of a competent national authority, or an external body (e.g., an Ombudsman).

Another important limitation of Article 11 relates to its personal scope, namely the exclusion of persons performing platform work who are also business users as defined in Article 2, point (1), of Regulation (EU) 2019/1150 (Article 11 (5)). This exception is remarkable given that the provision applies also to those performing platform work without employment status. The category of a genuinely self-employed platform worker tends to overlap with the definition of a business user.[39][40] Thus, this exclusion may render the protection effectively limited to platform workers with an employment status.

### *2.5. Safety and health*

'Promoting transparency, fairness, human oversight, *safety* and accountability in algorithmic management in platform work' is one of the goals of the Directive expressed in Article 1 (1) (b) PWD. Ensuring the safety of platform work has gained prominence in the final instrument compared to its previous versions. Not only has it been declared as one of the Directive's objectives,[41] but also a separate article, i.e., Article 12, has been devoted to it.[42] This can be read as a firm recognition of the algorithmic management as a factor aggravating OSH-related risks of platform workers. Research provides ample evidence on how automated monitoring and decision-making systems generate psychosocial risks.[43] Automated distribution of tasks, issuing instructions for work performance, as well as continuous and intrusive monitoring, have salient health and safety

[38] REPORT on the proposal for a directive of the European Parliament and of the Council on improving working conditions in platform work (COM(2021)0762 – C9-0454/2021 – 2021/0414(COD)), 21.12.2022, available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0301_EN.html.

[39] Recital 65 PWD

[40] Article 2 (1) of the Platform to Business Regulation defines a business user as 'any private individual acting in a commercial or professional capacity who, or any legal person which, through online intermediation services offers goods or services to consumers for purposes relating to its trade, business, craft or profession'

[41] The original proposal has declared only transparency, fairness, and accountability in algorithmic management as the Directive's objectives.

[42] In the original Proposal, this provision has been part of the provision on human monitoring of automated systems (previously Article 8).

[43] See e.g., BÉRASTÉGUI, Pierre. *Exposure to psychosocial risk factors in the gig economy: a systematic review*, Brussels, ETUI, 2021. Available at bit.ly/4049gJ9.

implications. Incentivising platform workers through the system of nudges and penalties to increase their work intensity and pace, often expecting them to surpass their limits, is a classical example of how algorithms put platform workers at risk.[44]

Safeguards laid down in Article 12 PWD should protect platform workers from an algorithmically-driven qualitative and quantitative overload, whereby what is required surpasses their expertise and abilities. Under this provision, platforms shall carry out a risk evaluation concerning the impact of algorithmic management on workers' safety and health. In particular, possible risks of work-related accidents, and psychosocial and ergonomic risks should be considered. Moreover, platforms shall evaluate whether the safeguards in place are adequate for the identified risks. They shall also introduce appropriate preventive and protective measures. This is an important step towards responsabilisation of platforms for the safety and health of people working through them. Importantly, for the first time in the EU legislation, explicit reference has been made to psychosocial risks, including undue pressure at work.[45] Since the exertion of pressure on workers as an indirect way of optimising their performance is driving the core model of the platform economy, it will be intriguing to see how this provision is applied and mobilised in courts and, in particular, what will be construed as 'undue pressure'. This provision certainly provides ample ground for litigation.[46]

Article 12 PWD does not go so far as to require platforms to 'program' OSH compliance into its algorithm, as could have been the case. It would be conceivable to require platforms to put in place mechanisms ensuring respect for working time regulations, among others. Lack of such an obligation is a missed opportunity for ensuring that algorithmic management not only does not exert detrimental effects on workers' safety and health but also enhances it by enhancing the enforcement of the OSH regulations.

Moreover, the provision does not contain any formal, procedural requirements regarding the risk assessment process. It does not specify the regularity with which it should be conducted, or the exact information to be provided. In particular, it is not clear whether the risk management should be conducted periodically, upon request, or before the introduction of the system. Article 12 of the Directive is silent also on the involvement of competent authorities, such as labour inspectorates, in obtaining access to information on OSH compliance of the algorithmic systems. The vagueness of this provision can undermine the effectiveness of the due diligence process platforms have to conduct. While Article 12 of the Directive mandates 'effective information and consultation and the participation of platform workers and/or their representatives', it could have gone further and, taking inspiration from the OSH Framework Directive. The latter instrument provided platform workers and their representatives also the right

[44] European Agency for Safety and Health at Work. Digital platform work and occupational safety and health: a review. 2021. Available at bit.ly/4dK7obx

[45] CEFALIELLO, Aude. An Occupational Health and Safety Perspective on EU Initiatives to Regulate Platform Work: Patching up Gaps or Structural Game Changers? Journal of work health and safety regulation. 2023 (1), 117-137.

[46] Ibid.

to appeal to national authorities responsible for OSH if they consider that the measures adopted at work do not protect them adequately (Article 11(6) Directive 89/391/EEC). As the provision stands now, there is no dedicated procedure provided for platform workers whose rights under Article 12 are not complied with.

It should be recalled, however, that OSH risks in platform work are multi-layered, as they do not stem only from algorithmic management.[47] They are related to the very nature of platform work, i.e., to precariousness, work fragmentation, instability, and isolation. Some risks, e.g., societal anxiety or even depression, may be related to mundane, repetitive tasks platform workers have to perform, and their perception of 'purposefulness'.[48] In other words, algorithmic management is an important factor triggering health and safety risks of platform workers, but by no means not the only one. Thus, even if Article 12 PWD is a welcome and significant step forward, it is far from ensuring comprehensive OSH protection. It cannot be seen as a panacea to all safety and health risks faced by platform workers.

Protecting workers' safety and health is even more complex in the configurations involving the client(s), to whom OSH risks can be attributed. A case in point could be the provision of inadequate equipment for a platform-mediated cleaner, or a pressure exerted by the client to drive faster. Article 12 (5) provides that Member States shall ensure platforms take preventive measures 'to ensure safety and health of platform workers, including from violence and harassment'. This provision stands out as the only one that seeks to address the duties of the platform in view of other party that may pose the risk to platform workers' health and safety. The Directive could have taken more account of these complex multi-partite constellations, furhter clarifying the responsibilities of platforms in this context.

The core problem regarding their inadequate protection relates to the personal scope of this provision, which is restricted to platform workers with an employment status. Platforms were therefore obliged to comply with a range of obligations under that Directive, as long as platform workers fall under the definition of a worker. The Platform Work Directive does not overcome the limitation of the OSH Framework Directive, which similarly applies only to employees. The (mis)classification of platform workers as independent contractors, and the persisting lack of clarity regarding the employment categorisation, is a key factor 'diluting' OSH responsibilities, challenging the applicability of OSH safeguards in the platform work context.[49] This is a critical issue, considering that all people performing platform work, irrespective of their employment contract, are exposed to physical

[47] Ibid, p. 221.

[48] This is experienced especially by platform workers on crowdwork platforms. According to some studies, approximately 50% of crowdworkers on Amazon Mechanical Turk experience clinical levels of social anxiety, while the figure for the general population lies at around 7-8%. See BÉRASTÉGUI, Pierre. *Exposure to psychosocial risk factors in the gig economy: a systematic review*, Brussels, ETUI, 2021, p. 16.

[49] LYKKE NIELSEN Mette, SLOTH LAURSEN Cæcilie, and DYREBORG Johnny. Who takes care of safety and health among young workers? Responsibilization of OSH in the platform economy. Safety Science Volume 149, 2022, 105674.

and psychosocial risk factors related to platform work. While the issue of misclassification is sought to be addressed by the other set of provisions of the Directive,[50] not all people performing platform work will benefit from the employment presumption and will continue to be exposed to high algorithmic-driven OSH risks. Neither will they be able to enjoy collective representation in matters related to the protection of their safety and health.

### *2.6. Information and consultation*

The last three provisions in Chapter III of the Platform Work Directive are dedicated to collective representation. While vast entitlements have been afforded to workers' representatives in a range of provisions in Chapter III[51] and other chapters of the Directive,[52] Articles 13-15 PWD introduce important specifications and complement the other provisions. First, Article 13 (2) clarifies that Member States shall ensure that information and consultation of workers' representatives by digital labour platforms also covers decisions likely to lead to the introduction of or to substantial changes in the use of automated monitoring or decision-making systems. Moreover, under Article 13 (3), the platform workers' representatives are entitled to the assistance of an expert of their choice, should it be necessary for them to investigate the matter that is the subject of information and consultation and issue an opinion. The expenses for the expert shall be borne by the digital labour platform if a platform has over 250 workers in a given Member State, provided that conditions of proportionality are met. Further, Article 14 stipulates that platform workers who have no representatives should be directly informed by the platform of decisions likely to lead to the introduction of or substantial changes in the use of algorithmic management systems. The information shall be provided in writing and be presented in a transparent intelligible and easily accessible form. Finally, Article 15 introduces constraints for representatives of persons performing platform work who do not enjoy employment status. Their access to information is limited strictly to their action on behalf of those providing services through platforms with regard to the protection of their personal data. This restriction stems from the legal basis, i.e., Article 16 TFEU.

[50] Articles 4-6 of the Platform Work Directive concerning the employment status.

[51] To name a few examples analysed above, workers' representatives have the right to obtain the data protection impact assessment under Article 8 (2), the right to be informed about the use of automated monitoring or decision-making systems under Article 9, and the right to be involved in human oversight of algorithmic systems under Article 10 (1).

[52] Article 19 provides that Member States shall ensure that representatives of persons performing platform work may engage in any procedure to enforce any of the rights or obligations arising from the Directive, including its algorithmic management provisions. Moreover, Article 28 clarifies that Member States are allowed to provide for more specific rules to ensure the protection of the rights and freedoms in respect of the processing of persons performing platform work's personal data under Articles 9, 10 and 11. They may also allow the social partners to maintain, negotiate, conclude and enforce collective agreements, in accordance with national law or practice, which, while respecting the overall protection of platform workers, establish arrangements concerning platform work which differ from those referred to in Articles 12 and 13.

## 3. Discussion and conclusion

Chapter III of the Platform Work Directive has been carefully drafted to respond, in a tailored and targeted way, to the most pertinent risks posed or aggravated by automated monitoring and automated decision-making systems used by digital labour platforms. It aptly addresses some of the key areas of concern, including insufficient data protection, risks for safety and health at work, lack of transparency, and adverse impact on working conditions. A comprehensive set of 'algorithmic rights' is provided to all people performing platform work, regardless of their employment status.

The new legal act decisively strengthens the existing protection under the General Data Protection Regulation and other instruments, including the OSH Framework Directive and the Directives in the field of anti-discrimination. The Platform Work Directive provides for a much more advanced set of protections than originally proposed by the Commission in December 2021, taking account of a vast bulk of the criticism that had been leveraged against the previous version by experts and activists. The main point of concern relates to the sectoral limitation of this instrument. In view of the 'spillover' of algorithmic management from digital platforms to other, more traditional segments of the labour market, ensuring broader protection against algorithmic risks is indeed a key regulatory challenge ahead.[53] At the same time, it should be acknowledged that the sectoral focus of the Platform Work Directive allowed it to 'shift away from 'risk' in the abstract',[54] and identify it in a possibly most concrete manner, which would not be possible with a more general, omnibus kind of legislative intervention.

Despite all the merits of the Platform Work Directive, the analysis presented in this paper has discussed a range of limitations of its algorithmic management provisions. In particular, it identified several exceptions in the (otherwise broad) personal and material scope of the provisions set out in Chapter III of the Directive, which are often overlooked in the discussion in literature. The three most critical ones concern the exclusion of business users under the Platform-to-Business (P2B) Regulation from the provisions on human review under Article 11 PWD, the limitation of the protection against OSH-related risks caused by algorithmic management under Article 12 to platform workers with employment status, and the exclusion of persons undergoing the recruitment and selection procedure from the human oversight provisions under Article 10.

Another key criticism formulated in the present contribution concerns the fragmentation of rights with regard to various categories of automated decision-making and automated monitoring systems. For instance, the impact assessment under Article 10 (3) PWD does not concern the risk of discrimination resulting from decisions taken by

[53] POTOCKA-SIONEK, Nastazja, ALOISI, Antonio. The 'spillover effect' of algorithmic management and how (not) to tame it. In Vandaele K. and Rainone S. (eds), *The Elgar companion to regulating platform work. Insights from the food delivery sector. Cheltenham*, Edward Elgar, 2024, forthcoming. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4738680.

[54] KELLY-LYTH, Aislinn and THOMAS, Anna. Algorithmic management: Assessing the impacts of AI at work. *European Labour Law Journal*, *14*(2), 2023, p. 251.

those systems. Another example is the limitation of transparency rights under Article 9 (1) (c) (iv) PWD to grounds on which automated systems refused to pay but would not cover a decision involving a change of the payment rate.

Besides the fragmentation of rights under the Directive, the paper drew attention to the lack of strong competencies of national authorities to intervene in the case of the algorithmically-driven infringement of workers' rights. Platforms are obliged to prevent the occurrence of such violations in the future and to compensate for the algorithmic harm, but they are entirely free to choose how they will achieve it. National authorities have merely information rights as regards the outcome of human oversight and review processes.

The effectiveness of the Directive's algorithmic management provisions will ultimately depend on the enforcement of this instrument by the Member States and its judicial interpretation. To date, only a few EU countries have introduced regulations on algorithmic management at work, including in platform work.[55] In most cases, these interventions represent a piecemeal approach, addressing only a fraction of algorithmic risks or merely reiterating existing protections under non-discrimination and data protection law.[56] In Spain, for instance, Article 64(4)(d) of the Workers' Statute provides important, collective safeguards against algorithmic risks by ensuring that works councils have the right to information about the parameters, rules, and instructions upon which algorithms are based, as long as they are used for decision-making practices that may affect working conditions, access to employment or maintenance of employment, including profiling.[57] The provision largely mirrors Article 9 of the Platform Work Directive. Still, it does not amount to an obligation to negotiate the algorithm with the workers' representation, and neither does it provide redress mechanisms. Therefore, further legislative intervention is needed. The reform of the Croatian labour law, effective as of January 2024, stands out as a good example of a regulation which, albeit platform-specific, provides for a comprehensive set of safeguards against algorithmic management risks, in some aspects surpassing the protection provided for under the Platform Work Directive.[58]

[55] For an overview, see LITARDI Chiara, ADĂSCĂLIȚEI Dragoș, and WIDERA Sarah. Anticipating and managing the impact of change. Regulatory responses to algorithmic management in the EU. 21 May 2024. Available at https://www.eurofound.europa.eu/en/resources/article/2024/regulatory-responses-algorithmic-management-eu.

[56] POTOCKA-SIONEK, Nastazja, ALOISI, Antonio. The 'spillover effect' of algorithmic management and how (not) to tame it. In Vandaele K. and Rainone S. (eds), *The Elgar companion to regulating platform work. Insights from the food delivery sector. Cheltenham*, Edward Elgar, 2024, forthcoming.

[57] TODOLÍ-SIGNES, Adrian. Spanish riders law and the right to be informed about the algorithm. European Labour Law Journal, 12(3), 2021, 399-402.

[58] The unofficial consolidated text of the Labour Act is available at https://uznr.mrms.hr/wp-content/uploads/labour-act.pdf. An important addition to the obligations stemming from the Platform Work Directive is the obligatory appointment of an authorised person to supervise platform workers' safety and workload, as well as to carry out a review of automated decisions made in the automated management system and decide on them at the request of the worker. Moreover, after reviewing the decision, platform worker should be authorised to an expert statement of the decision and to decide upon the review of the decision.

## 4. Bibliography

ABRAHA, Halefom. Regulating algorithmic employment decisions through data protection law. *European Labour Law Journal*, *14*(2), 2023.ADAMS-PRASSL, Jeremias. The Challenges of Management by Algorithm: Exploring Individual and Collective Aspects. Gyulavári, Tamás , and Emanuele Menegatti, ed. Decent Work in the Digital Age: European and Comparative Perspectives. Oxford: Hart Publishing, 2022. Bloomsbury Collections.BAIOCCO Sara, FERNANDEZ-MACÍAS Enrique, RANI Uma, PESOLE Annarosa.

– The Algorithmic Management of work and its implications in different contexts. *Background Paper Series of the Joint EU-ILO Project "Building Partnerships on the Future of Work"*. Geneva, 2022.

BÉRASTÉGUI, Pierre. *Exposure to psychosocial risk factors in the gig economy: a systematic review*, Brussels, ETUI, 2021. Available at bit.ly/4049gJ9.

CEFALIELLO, Aude. An Occupational Health and Safety Perspective on EU Initiatives to Regulate Platform Work: Patching up Gaps or Structural Game Changers? *Journal of work health and safety regulation.* 2023 (1); 117−137.

DUBAL, Veena. On algorithmic wage discrimination. *Columbia Law Review*, 123(7), 2023, 1929-1992.

DUGGAN, James, SHERMAN Ultan, CARBERY Ronan, and MCDONNELL, Anthony.

– Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. Human Resource Management Journal, 30(1), 114-132.

European Agency for Safety and Health at Work. *Digital platform work and occupational safety and health: a review.* 2021. Available at https://osha.europa.eu/sites/default/files/2021-11/OSH_policies_online_platform_economy.pdf

FIERS, Floor. Inequality and discrimination in the online labor market: A scoping review. *New Media & Society*, *25*(12), 2023, 3714-3734

GUGLIELMETTI Mario. Automated work and workers' rights: platform work and AI work management systems. In Artificial intelligence, labour and society, Aída Ponce Del Castillo (ed.), European Trade Union Institute (ETUI), Brussels, 2024. Available at bit.ly/4e1z641

HIESSL, Christina. Case Law on Algorithmic Management at the Workplace: Cross-European Comparative Analysis and Tentative Conclusions. 2023. Available at https://ssrn.com/abstract=3982735

HIESSL, Christina. Algorithmic Management in the Workplace: Taking Stock of Case Law and Litigation in Europe. Hungarian Labour Law e-Journal. 2022/2. Available at https://hllj.hu/letolt/2022_2_a/01_ChHiessl_hllj_uj_2022_2.pdf.

KELLY-LYTH, Aislinn and THOMAS, Anna. Algorithmic management: Assessing the impacts of AI at work. *Eusropean Labour Law Journal*, *14*(2), 2023, 230-252.

LEE, Min Kyung, KUSBIT, Daniel, METSKY, Evan, and DABBISH, Laura. Working with machines: The impact of algorithmic, data-driven management on human workers. Proceedings of the 33rd Annual ACM SIGCHI Conference, Seoul, South Korea, 2015. LITARDI Chiara, ADĂSCĂLIȚEI Dragoș, and WIDERA Sarah. Anticipating and managing the impact of change. Regulatory responses to algorithmic management in the EU. 21 May 2024. Available at https://www.eurofound.europa.eu/en/resources/article/2024/regulatory-responses-algorithmic-management-eu. LYKKE NIELSEN Mette, SLOTH LAURSEN Cæcilie, and DYREBORG Johnny. Who takes care of safety and health among young workers? Responsibilization of OSH in the platform economy. Safety Science Volume 149, 2022, 105674.

MARTINDALE, Nicholas, WOOD, Alex J., and BURCHELL, Brendan. What do platform workers in the UK gig economy want? *British Journal of Industrial Relations* 62(3), 2024, 542-567.

MATEESCU, Alexandra and NGUYEN, Aiha. Explainer: Algorithmic management in the workplace. 2019. Available at https://datasociety.net/wpcontent/uploads/2019/02/DS_Algorithmic _Management_E xplainer.pdf

OGUNDE, Fife. Algorithmic management of platform workers: An examination of the Canadian and European approaches to regulation. *European Labour Law Journal*, 0(0), 2024. Available at https://journals-sagepub-com.eui.idm.oclc.org/doi/epub/10.1177/20319525241239632 PONCE DEL CASTILLO, Aída and NARANJO, Diego. *Regulating algorithmic management.* An assessment of the EC's draft Directive on improving working conditions in platform work. ETUI Policy Brief 2022.08. Available at https://bit.ly/4f5y9sj

PONCE DEL CASTILLO, Aída. Regulating algorithmic management in the Platform Work Directive: correcting risky deviations. Global Workplace Law & Policy. Available at bit.ly/3Y63EeF

POTOCKA-SIONEK, Nastazja and ALOISI, Antonio. The 'spillover effect' of algorithmic management and how (not) to tame it. In Vandaele K. and Rainone S. (eds), *The Elgar companion to regulating platform work. Insights from the food delivery sector. Cheltenham*, Edward Elgar, 2024, forthcoming. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4738680.

RAINONE, Silvia and ALOISI, Antonio. The EU Platform Work Directive What's new, what's missing, what's next? ETUI Policy Brief, 2024.06, August. Available at bit. ly/4eIMYRA

RENAN BARZILAY, Arianne. The Technologies of Discrimination: How Platforms Cultivate Gender Inequality. *The Law & Ethics of Human Rights*, vol. 13, no. 2, 2019, pp. 179-202.

SULLIVAN Rick, VEEN Alex, and RIEMER Kai. Furthering engaged algorithmic management research: Surfacing foundational positions through a hermeneutic literature analysis. Information and Organization Volume 34, Issue 4, December 2024, 100528.

TODOLÍ-SIGNES, Adrián. Spanish riders law and the right to be informed about the algorithm. European Labour Law Journal, 12(3), 2021, 399-402.

VEALE, Michael, SILBERMAN, Michael 'Six', and BINNS, Reuben. Fortifying the algorithmic management provisions in the proposed Platform Work Directive. *European Labour Law Journal*, 14(2), 2023.

WOOD, Alex J. Algorithmic management consequences for work organisation and working conditions, JRC Working Papers Series on Labour, Education and Technology 2021/07, 2021, European Commission.

# La videovigilancia en el trabajo en tiempos de inteligencia artificial

## Video surveillance at the workplace in times of artificial intelligence

Miguel Rodríguez-Piñero Royo
*Universidad de Sevilla*
ORCID ID: 0000-0001-7926-6175

doi: 10.20318/labos.2024.9032

*Resumen:* Este trabajo analiza los efectos de la integración de sistemas de IA en la videovigilancia en las empresas, y qué regulación se aplica cuando esto ocurre. Se estudia la aplicación conjunta de las reglas sobre videovigilancia del Derecho digital del trabajo con las del Derecho Algorítmico. La conclusión es que dispone de reglas bastante completas para ordenar estos sistemas, sin que se detecten grandes contradicciones entre ambos sectores. Estos dispositivos y la analítica de imágenes incrementan los riesgos de vulneración de derechos para los trabajadores, pero también proporcionan a los empleadores un importante instrumento para la mejora de la salud y la seguridad, la toma de decisiones y el cumplimiento normativo, por lo que debe avanzarse en su regulación.

*Palabras clave:* Videovigilancia, Inteligencia Artificial, poderes empresariales, cumplimiento normativo, derechos de los trabajadores.

*Abstract:* This paper analyzes the effects of integrating AI systems into workplace surveillance and the applicable regulations when this occurs. It examines the joint application of digital labor law surveillance rules with those of Algorithmic Law. The conclusion is that there are fairly comprehensive rules to regulate these surveillance systems, without significant contradictions between the two sectors. These devices and image analytics increase the risk of rights violations for workers, but also provide employers with an important tool for improving health and safety, decision-making, and regulatory compliance. Therefore, progress must be made in their regulation.

*Keywords:* Videosurveillance, Artificial Intelligence, managerial powers, compliance, workers' rights.

## 1. Presentación: control, tecnología y derecho

El Derecho del Trabajo que hoy conocemos es, todavía, el del trabajado asalariado y por cuenta ajena, al que calificamos como subordinado porque la nota que lo caracteriza, frente a otras formas de empleo retribuido, es el sometimiento de la persona que lo desempeña a otra. La evolución del ordenamiento jurídico ha llevado a que esta dependencia sea una estrictamente contractual, derivada de la voluntad de la persona que acepta formar parte de un vínculo obligacional, lejos ya de compromisos personales o de estatus de clase. La dependencia supone, como es bien sabido, el sometimiento a tres poderes empresariales, que generan derechos para los empleadores y correlativas obligaciones para quiénes trabajan para ellos: poder de dirección, poder disciplinario y poder de control. Del último de ellos me ocuparé en este trabajo.

Se trata de un poder legalmente reconocido, justificado y finalista en la medida en que se ejercita en relación con una serie limitada de finalidades, amén de otras restricciones derivadas por la forma en que afecta a derechos fundamentales de la persona, tanto los generalmente conocidos como "inespecíficos", como otros que podemos considerar específicos del trabajador, que es titular de una especie de "intimidad laboral" que ha estado reconocida desde un primer momento, antes incluso de que se pensara en términos de derechos de éstos. Es un poder marcadamente pro empresario, para el cuidado de sus intereses en la relación de trabajo, aunque no exclusivamente desde el momento en que se ejercita para verificar el cumplimiento de determinados derechos de los trabajadores.

Estos tres poderes, por más que como laboralistas nos hayamos acostumbrados a ellos hasta normalizarnos, no dejan de ser anómalos en un contexto contractual, con un Derecho de obligaciones construido sobre modelos completamente diferentes. La función histórica del Derecho del Trabajo, en el marco del Derecho privado, ha sido la de establecer un marco normativo que permita su ejercicio de acuerdo con unos estándares de dignidad y de derechos.

Poder de dirección y poder de control comparten estar directamente condicionados por el estatus quo tecnológico del momento en que se ejercitan. Es la tecnología la que determina cómo se trabaja, y cómo se supervisa lo que se trabaja. De ahí que su ordenación sea en gran medida la de los medios técnicos utilizados por la empresa, cuyos avances generan problemas crónicos de obsolescencia regulatoria. El impacto de la innovación no se produce sólo en el ejercicio de las facultades de dirección y supervisión, pero no cabe duda de que incide especialmente en éstas.

Las últimas revoluciones tecnológicas, digitalización e inteligencia artificial (IA), han tenido un especial impacto en este ámbito, al haber generado nuevos instrumentos y haber extendido los ya existentes. Puede decirse incluso que, en las últimas décadas, y como consecuencia de ellas, se está produciendo un cambio en el peso específico de cada uno de ellos en la relación de trabajo: los medios disponibles han permitido una mayor autonomía en la prestación de trabajo, con lo que la relevancia de la dirección se reduce, a la vez que se han incrementado las posibilidades de controlar, haciendo que la

supervisión resulte más importante. En los centros de trabajo del siglo XXI la autonomía en el trabajo no se ve acompañada de un aligeramiento del control, sino más bien de lo contrario, siendo éste a lo que las empresas han recurrido para mantener el gobierno de la organización.

Tradicionalmente entre los indicios de laboralidad manejados por nuestros tribunales escaseaban los que tenían que ver con el control de la actividad del trabajador, en detrimento de aquellos vinculados con el ejercicio de los poderes directivos. El desafío del trabajo digital, en particular el de las nuevas formas de empleo generadas por éste, está haciendo que el foco vaya dirigiéndose también a este aspecto, alineando esta tarea con la realidad del trabajo en este siglo.

El cambio en los soportes que facilitan el control empresarial se ha percibido por lo general en términos de riesgo para los trabajadores, conceptuándose la idea de la "empresa panóptica" (MERCADER UGUINA), como un prototipo ucrónico en el que los derechos de éstos podían verse sistemáticamente limitados. Y es cierto que tanto los espacios tradicionales de intimidad como las expectativas de privacidad se ven afectados por las capacidades ampliadas de supervisión, todo ello en un contexto jurídico de mayor sensibilidad hacia los derechos vinculados con la protección de datos y la intimidad. Sin embargo, y de una manera ciertamente paradójica, el control de la empresa está mutando para desarrollar otra de sus facetas, la de la garantía del cumplimiento normativo en lo laboral. Esta doble naturaleza no es nueva, pero sí se ha visto potenciada con figuras tales como el registro de jornada, originalmente diseñado para asegurar el cumplimiento de las obligaciones laborales, hoy orientado también hacia la evitación de abusos en el tiempo de trabajo. Esto debe ser tenido en cuenta, a mi juicio, cuando se analizan los instrumentos de vigilancia en el nuevo entorno tecnológico, para aprovechar al máximo las posibilidades que se generan.

En este trabajo voy a analizar un mecanismo particular de control empresarial, la videovigilancia, cuyo análisis resulta especialmente interesante por una serie de motivos: es una forma con una larga tradición, que precede la existencia de soportes tecnológicos que lo faciliten (un supervisor en un centro de trabajo está vigilando la actividad de los empleados utilizando su propia vista); ha experimentado sucesivas innovaciones técnicas, que han tenido el efecto de hacerla ubicua en el siglo XXI; su régimen jurídico ha ido evolucionando, de acuerdo con el contexto regulatorio de la tecnología en cada momento; y, finalmente, puede verse especialmente afectada por el desarrollo de la IA, de la que nos estamos ocupando en este número monográfico.

La hipótesis de partida de este estudio es que se está produciendo la construcción de un Derecho algorítmico del trabajo, constituido por un conjunto de respuestas jurídicas a los problemas que se percibe puede generar la utilización de sistemas de inteligencia artificial en el ámbito laboral. A la vez, la videovigilancia se está transformando por la integración de estos sistemas, de tal modo que su utilidad, alcance y efectos van a cambiar radicalmente. Las reglas jurídicas ya aplicables deberán actualizarse para adecuarse a estos nuevos usos, a la vez que se ajustan a los principios e instrumentos del Derecho algorítmico.

## 2. La videovigilancia en el derecho del trabajo anterior a la inteligencia artifical

Como ya se ha dicho, la videovigilancia es anterior a la digitalización, surgiendo en un entorno analógico, lo que hizo que originalmente tuviera poco impacto en la práctica, por su escaso alcance y limitada utilización. Su regulación estuvo condicionada por este estatus quo tecnológico, lo que se tradujo en un tratamiento legal limitado, cuando existía, y la aplicación de la normativa general sobre intimidad de la persona trabajadora y poderes empresariales. Tampoco era frecuente su tratamiento en los convenios colectivos, lo que dejaba el protagonismo a la jurisprudencia, tanto laboral como constitucional, que se realizaba aplicando la regulación del derecho a la intimidad e importando reglas de otros sectores del ordenamiento (como el Derecho procesal en torno a la validez de las grabaciones como prueba).

Con el tiempo una serie de cambios afectó a este mecanismo de control, algunos técnicos y otros de distinta naturaleza. Así, uno de los factores que más cambió su utilización fue el abaratamiento de los soportes técnicos utilizados, tanto para la obtención de imágenes como para su almacenamiento. Esto permitió un uso extensivo de la videovigilancia. También ayudó la miniaturización de las cámaras, que permitió su ubicación en todo tipo de entornos, así como su inclusión en otros soportes (como ordenadores y monitores), lo que llevó una especie de invisibilidad, al perderse la consciencia de estar siendo monitorizado. No en vano una de las principales medidas de control que se le impone es, precisamente, la señalética para avisar de su presencia.

La digitalización alteró profundamente su utilización. Al pasarse las imágenes obtenidos a un formato digital se facilitó su almacenamiento, tratamiento, transferencia y manipulación. Convertida en un dato, la imagen de las personas se hace merecedora de tutela por la legislación de protección de datos.

Hoy está presente en todo tipo de entornos laborales, así como en aquellos otros que, no siéndolo en sentido estricto, pueden exponer a las personas que trabajan al control de su imagen, como las vías públicas durante los desplazamientos causados por la actividad profesional.

Estos avances han cambiado radicalmente la percepción de riesgos asociados a su presencia, lo que ha impulsado el desarrollo de su régimen jurídico. Éste se ha construido sobre jurisprudencia anterior, y ha combinado reglas generales y especiales. En particular se ha definido un nuevo derecho fundamental laboral, encuadrado en los derechos digitales, que es el derecho a la tutela frente a la videovigilancia abusiva o excesiva (no frente al control de imágenes en sí mismo). De esta materia, la Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital (2023/C 23/01) afirma que "*nos comprometemos a (…) la protección frente a una vigilancia ilegal e injustificada*". En España la Carta de Derechos Digitales dispone que "*en los entornos digitales y el teletrabajo las personas trabajadoras del sector público o privado tienen derecho con arreglo a la normativa vigente, a la protección de sus derechos a la intimidad personal y familiar, el honor, la propia imagen, la protección de datos y el secreto de las comunicaciones frente al uso de dispositivos de videovigilancia*".

Este derecho, que es autónomo frente a los genéricos de intimidad y protección de datos, es la base para un régimen jurídico diferenciado, que en nuestro país incluye dos preceptos que no llegan a ser monográficos, aunque sí aportan mandatos para definir éste. Por un lado, el artículo 20 bis del Texto Refundido del Estatuto de los Trabajadores (TRET), que contempla los derechos de los trabajadores a la intimidad en relación con el entorno digital y a la desconexión, entre los que se incluye el de la intimidad frente al uso de dispositivos de videovigilancia. El precepto estatutario se remite a los términos establecidos en la legislación vigente en materia de protección de datos personales y garantía de los derechos digitales en cuanto al ejercicio de este derecho, para lo que disponemos, como segundo referente normativo, del artículo 89 de la Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales, que reconoce el derecho a la intimidad frente al uso de dispositivos de videovigilancia y de grabación de sonidos en el lugar de trabajo.

La combinación de estos preceptos con las construcciones jurisprudenciales anteriores resulta en un régimen caracterizado por siguientes notas:

— Relevancia del dispositivo, de tal modo que esta regulación se aplica a un tipo determinado de éstos, los que recogen imágenes.
— Relevancia del tipo de dato, dado que se trata de reglas específicas para las imágenes.
— Legitimidad del uso de estos dispositivos por parte de las empresas, pero limitado al ejercicio de las funciones de control de personas trabajadoras que emplean, y en relación con la actividad productiva y la seguridad de las personas.
— Sometimiento de este control mediante videovigilancia al marco legal y a los límites inherentes al mismo. Esto incluye la legislación de protección de datos
— Deber de información a las personas trabajadoras que van a ser objeto de esta medida de control; esta información deberá suministrarse con carácter previo, y de forma expresa, clara y concisa.
— Deber de información a los representantes de la plantilla, en caso de estar constituida ésta, en los mismos términos que a los trabajadores individuales.
— Aplicación de criterios de necesidad, idoneidad y proporcionalidad en cuanto a la utilización de este control.
— Deber de minimización, que supone (AEPD) que se valore si realmente es necesaria la instalación de la videovigilancia, o si el fin perseguido se puede alcanzar de otra forma. Además, cuando realice la instalación, que se tenga en cuenta la proporcionalidad en función del número de cámaras, tipo de las mismas y la opción de utilizar "máscaras de privacidad".
— Respeto a las expectativas de intimidad, de tal modo que, si existen motivos justificados para considerar que no se va a ser objeto de este control, su utilización resultará ilegítima.

— Respeto a los espacios de intimidad, en los que no se admite su uso, y que incluyen los lugares destinados al descanso o esparcimiento, como vestuarios, aseos, comedores y análogos.
— Reconocimiento del papel de los representantes de los trabajadores, aunque desde nuestra perspectiva nos parece un rol bastante discreto.

Por su propia naturaleza, las reglas propias de la videovigilancia se superponían frecuentemente con otros contenidos del Derecho digital, desde el punto de vista de la utilización de las imágenes obtenidas. Es frecuente que estos dispositivos sean instrumentales para el establecimiento de controles biométricos, que están admitidos pero sujetos a exigencias rigurosas, por su carácter intrusivo y porque los riesgos vinculados a este tipo de controles van más allá de los ya identificados para los de imagen. La videovigilancia entra igualmente dentro del ámbito de aplicación de la normativa de protección de datos, en cuanto la imagen es considerada legalmente como un dato, y estos sistemas permiten el almacenamiento y la transmisión de éstas. Es posible que este control se realice mediante dispositivos empresariales puestos a disposición de los trabajadores, como las cámaras de los ordenadores portátiles. Si además estos se utilizan en relaciones de teletrabajo en el domicilio de la persona trabajadora, se le aplica el conjunto normativo especial que regula esta forma de empleo.

Comparte la regulación de la videovigilancia con otras formas del Derecho Digital la presencia de una gran diversidad de elementos normativos, ya que junto a las normas legales encontramos otras propias del soft-law, cláusulas convencionales, guías de uso, cláusulas contractuales y normativas internas de las empresas.

## 3. Del Derecho digital al Derecho algorítmico

La aparición de los sistemas de IA está suponiendo una verdadera revolución, que ha impulsado una nueva fase en la evolución del tratamiento de los instrumentos tecnológicos por el ordenamiento jurídico. En efecto, si realizamos un análisis histórico superficial del tratamiento de esta cuestión, podemos identificar diferentes fases o momentos, comenzando por regulaciones concretas para los mecanismos de control más generalizados (por ejemplo, para los registros en la persona del trabajador), para llegar a la utilización de los llamados "derechos inespecíficos" para garantizar ciertos niveles de protección a las personas. El desarrollo de la legislación de protección de datos afectó igualmente a los sistemas de vigilancia en la empresa. A finales del siglo pasado se manejó una categoría, la de "derechos on-line", que no tuvo mucho impacto en el Derecho positivo pero que sí sirvió para llamar la atención sobre los riesgos derivados de la utilización de mecanismos avanzados de control y de la generalización de las entonces conocidas como TIC. La fase siguiente, en la que todavía nos encontramos, es la del Derecho Digital del Trabajo, que dedica una parte importante de su regulación a la tutela de las personas frente a los mecanismos de supervisión basados en esta tecnología, videovigilancia incluida.

Ahora estamos asistiendo al surgimiento de un nuevo conjunto regulador, conocido generalmente como Derecho Algorítmico o Derecho de la IA, que tiene una vertiente laboral en cuanto regula la utilización de tales sistemas en la gestión de personas, incluyendo su supervisión y control. Esta rama no debe entenderse estrictamente como una nueva fase en la evolución del Derecho de la tecnología en el trabajo, sino como un desarrollo monográfico de una parte de éste, ya que supone una regulación propia para un tipo particular de instrumento empresarial que podría considerarse como parte de la tecnología digital, o que al menos se combina con ésta. Podríamos hablar, así, de que esta rama emergente como un "spin-off" del Derecho digital, centrado en la ordenación de los sistemas de IA que van a utilizarse en las empresas y en las administraciones públicas. Como tal derivación tienen un código genético compartido con éste, ya que comparte muchos de sus objetivos, principios e instrumentos. Tiene, aun así, señas de identidad propias, que lo diferencian y caracterizan como una regulación verdaderamente original.

Entre estas señas de identidad podemos señalar su construcción acelerada, ya que en un corto plazo se dispone ya de un conjunto de normas bastante extenso, resultado de una intervención multinivel que también lo caracteriza. Este desarrollo rápido es consecuencia de su carácter preventivo, no en el sentido de que se elaborara antes de que los sistemas de IA fueran una realidad en las empresas, sino que su construcción comenzó cuando surgió la preocupación por los potenciales riesgos que se atribuían a ésta IA. Esto es, que no se ha esperado a que se generaran los problemas para comenzar a diseñar sus soluciones, sino que éstas los han precedido, a partir de previsiones y expectativas razonables. En este sentido el Derecho algorítmico ha surgido con un marcado carácter académico y tecnocrático, puesto que está siendo elaborado por expertos de diversas disciplinas a partir de la constatación de sus efectos potenciales (o reales, como en el caso del trabajo en plataformas digitales), con el objetivo de adelantarse y anticiparse a éstos.

Esta es una importante diferencia con el Derecho Digital, al menos tal y como éste se había venido construyendo. Éste es igualmente de elaboración multinivel, y utiliza también una pluralidad de instrumentos reguladores. Ahora bien, éste ha tardado mucho más en desarrollarse, ya que sus grandes productos normativos se han aprobado cuando las consecuencias de la utilización de la tecnología digital en las empresas eran ya evidentes. Son muchas las reglas que codifican soluciones preexistentes, elaboradas por los tribunales internacionales o nacionales a partir de regulaciones sobre derechos fundamentales.

Su peculiar origen y su objeto material de regulación han dado lugar a otro rasgo característico del Derecho Algorítmico, su complicación y sofisticación técnicas. Vamos a encontrar regulaciones de realidades tecnológicas complejas y exigentes, cuya elaboración y aplicación exigen un alto nivel de conocimiento previo. A diferencia de las del Derecho Digital, sus normas son muchas más precisas, identificando con detalle su ámbito de aplicación y sus mandatos. Esto tiene que ver con tanto con la realidad material que regula como son su origen tecnocrático. Pero también con el hecho de que el Derecho Algorítmico, especialmente el de la Unión Europea, se ocupa de regular unos programas que van a desarrollarse, comprarse, exportarse. Pretende estandarizar unos

productos de nueva creación para facilitar su uso y circulación. El Reglamento de Inteligencia Artificial afirma en este sentido que su objetivo es "*mejorar el funcionamiento del mercado interior y 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 consagrados en la Carta, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA en la Unión así como prestar apoyo a la innovación*". Estamos ante una regulación que tutela derechos, pero que a la vez promociona un mercado y facilita el avance de la tecnología. Para ello establece, entre otras cosas, normas armonizadas para la introducción en el mercado, la puesta en servicio y la utilización de sistemas de IA en la Unión; así como medidas en apoyo de la innovación.

La regulación algorítmica es transversal, en el sentido de que se ocupa de unos sistemas que pueden adoptar distintos formatos y manejar datos de diferente naturaleza, dado que lo relevante para su categorización es un factor que opera en un plano diferente, el de los riesgos. Ser calificado con un mismo nivel de riesgo determina la aplicación de un régimen jurídico común; y este nivel se asigna en función del uso que se de al sistema de IA. El Derecho digital es, por el contrario, mucho más específico, ya que diferencia según el dispositivo de control que se utilice y el dato que se recopile, puesto que en la mayoría de los casos la intervención del empleador se hace con una misma función, monitorizar al empleado. Por poner un ejemplo, el Reglamento de Inteligencia Artificial maneja un concepto de sistema de identificación biométrica remota que incluye a todos los destinados a identificar a personas físicas sin su participación activa, con independencia de la tecnología, los procesos o los tipos de datos biométricos concretos que se usen

Finalmente, y si lo comparamos con el Derecho digital previo, el de la IA es algo más sentible a la dimensión colectiva de las relaciones laborales, a la que presta mayor atención. Se reconoce, en este sentido, el papel de los representantes de los trabajadores y las organizaciones sindicales, y se identifican los derechos sindicales como una realidad a tutelar.

## 4. El impacto de la IA en la videovigilancia

La IA se suma a los cambios tecnológicos experimentados por la videovigilancia en las últimas décadas, ya señalados previamente. Ésta ha mejorado tanto por los avances en el hardware, mediante incrementos de alcance y calidad de las imágenes, capacidad de tratamiento, minutarización y conectividad, como por el software. Y éste, a su vez, ha recibido mejoras tanto en los programas de gestión como con la introducción de la IA en su manejo. La IA se combina con una tecnología ya avanzada para incrementar las capacidades de monitarización, y esto ha llevado a que no se quede en esta función sino que va a servir también para apoyar la toma de decisiones por parte de las empresas, en ejercicio de su poder de dirección.

Esta combinación, que produce lo que en el lenguaje comercial de las empresas de seguridad se llama "cámaras inteligentes", está teniendo un gran impacto tanto por su generalización y extensión, muy rápida, como por los riesgos que se identifican en ella. Es la videovigilancia inteligente en algunas áreas, como la seguridad pública, la que ha movilizado muchos de los debates sobre los riesgos para los derechos fundamentales y las libertades públicas, y lo que ha dado lugar a regulaciones restrictivas, que encontramos en el mismo Reglamento de Inteligencia Artificial. Junto a ello se ha desarrollado la "analítica de imágenes", que consiste en la obtención de información por medio de herramientas de IA a partir de las imágenes obtenidas por un sistema de captación de éstas.

La videovigilancia inteligente está muy extendida, hablándose entonces de "videovigilancia masiva". Puede resultar invisible, no tanto por sus dimensiones o colocación, sino por su presencia generalizada en todos los ámbitos, que hace que se acabe por ignorarlas. En sentido contrario, el "sentimiento de vigilancia masiva", la consciencia de estar continuamente monitorizado, puede afectar negativamente a las personas, produciendo ansiedad y condicionando sus comportamientos.

La videovigilancia inteligente es instrumental para otras actuaciones, como el reconocimiento facial o la detección o deducción de emociones. El Reglamento de IA define los sistemas de identificación biométrica remota como aquellos destinados a identificar a personas físicas sin su participación activa, generalmente a distancia. Éstos pueden ser en tiempo real, si la recogida de los datos biométricos, la comparación y la identificación se producen de manera instantánea, e implican el uso de materiales en directo o casi en directo, como grabaciones de vídeo, generados por una cámara u otro dispositivo con funciones similares. En los sistemas en diferido también se utilizan imágenes o grabaciones de vídeo captadas por cámaras de televisión en circuito cerrado generados con anterioridad a la utilización del sistema en relación con las personas físicas afectadas.

En cuanto a la deducción de emociones de los trabajadores ésta puede basarse en las imágenes obtenidas mediante estos sistemas, tanto de expresiones faciales como de pautas de comportamiento y conductas concretas. No entraré en esta cuestión, dado que ésta será objeto de un análisis completo en otra colaboración a este número monográfico, por parte de una gran experta en este tema.

Las imágenes obtenidas mediante estos sistemas se utilizan para el desarrollo y el entrenamiento de los modelos de IA de uso general, en particular los grandes modelos de IA generativos, capaces ellos mismos de generar imágenes. Esta capacidad de generar imágenes con un alto grado de realismo y verosimilitud produce el riesgo de pruebas falsas, creadas o manipuladas, sobre las que basar decisiones empresariales, algo que con la videovigilancia digital resultaba mucho más complicado.

Se ha señalado igualmente que estos sistemas pueden extender su utilidad más allá del control de la realidad de la actividad laboral, para alcanzar otros aspectos como su cantidad y calidad, convirtiéndose en una fuente de información para la evaluación del desempeño. Una cámara inteligente controla automáticamente lo que se hace, cómo se hace, quién lo hace, en cuánto tiempo, etc.

La IA introducida en los aparatos de videovigilancia se utiliza combinación con otras tecnologías, como los drones y las cámaras "on board" en vehículos. Se diseñan para moverse en función de la información que estén recibiendo, siguiendo a personas y vehículos para tenerlos controlados de manera continuada. También se dotan de sensores para medir la temperatura y otros parámetros físicos y químicos del medio de trabajo. Pueden incluir herramientas de cómputo para determinar el número de personas presentes en un espacio determinado. Se activan o desactivan según se identifique la presencia de elementos predeterminados.

Además, en la medida en que se trata de un software que puede ser instruido fácilmente demuestra una gran capacidad de adaptación al usuario. Así, existen cámaras que descartan las imágenes de las mascotas en los domicilios dotados con estos sistemas de seguridad, o que se adaptan para la custodia a distancia de bebés.

La integración de la IA en la videovigilancia permite el análisis de las imágenes grabadas, su interpretación y la predicción de eventos, de manera inmediata. También hace posible el análisis en tiempo real de la información obtenida por las cámaras, y el aprendizaje automático a partir de ésta.

El sistema de IA inserto en la cámara identifica los elementos que aparecen en la grabación, lo que facilita la extracción de información de ésta, y así hace posible un control que con cantidades ingentes de imágenes resultaría imposible. La actividad de los responsables de seguridad se facilita enormemente, aunque con ello también el riesgo de intrusiones excesivas en la intimidad de las personas trabajadoras.

Las cámaras inteligentes no sólo captan imágenes, sino que detectan patrones y tendencias en la realidad observada, y esto genera el riesgo de control de emociones, como se ha dicho. También introduce el factor de los errores experimentados por la IA, que puede llegar a conclusiones equivocadas, experimentar alucinaciones o sufrir sesgos. De la misma manera, la IA puede concentrar la atención y la captación de imágenes en algunas personas, identificadas a partir de pautas constatadas o de predicciones, y esto genera un riesgo real de discriminación algorítmica.

Como resultado, la IA está transformando el uso que se hace de la videovigilancia en las empresas, permitiéndoles un nivel de control desconocido hasta ahora, tanto por la cantidad y calidad de la información visual que se recoge como por la posibilidad de obtener otra información a partir de ésta. Con las imágenes recogidas y tratadas se basan decisiones empresariales que nada tienen que ver con el cumplimiento de las obligaciones laborales por las personas empleadas por la entidad, sino en la búsqueda de eficiencia, la mejora de la productividad, el ahorro de costes, el incremento de la seguridad y otras finalidades, que no siempre resultan legítimas. Pensemos en la utilización de imágenes para predecir el comportamiento de los trabajadores frente a una huelga, unas elecciones sindicales o un proceso de certificación sindical (como efectivamente parece haber ocurrido en los Estados Unidos); o simplemente para calibrar el clima laboral o el impacto de decisiones empresariales que afectan a la plantilla.

Por todo ello se ha asociado su utilización con la aparición de nuevos riegos para los derechos de los trabajadores. Entre estos riesgos se pueden citar un uso extensivo y

sistemático del control de imágenes; la manipulación y generación de éstas; la posibilidad de alucinaciones y deducciones equivocadas; el control selectivo de ciertas personas o grupos de éstas, como consecuencia de sesgos que pueden suponer una verdadera discriminación; y el agotamiento del personal monitorizado.

Es igualmente cierto que la integración de la IA en los sistemas de control de imágenes proporciona a éstos nuevas utilidades, mejorando la gestión empresarial y la situación de los mismos trabajadores. Podemos pensar en la mejora en la seguridad de los centros de trabajo frente a la actuación de personas, empleadas o no, que puedan detectarse a tiempo o incluso con anticipación; la prevención de accidentes, especialmente incendios (ya se utiliza la videovigilancia inteligente para combatir los incendios forestales); el control sanitario (utilizado durante la pandemia, al poder detectarse la temperatura corporal y otros parámetros, así como detectar cuando una persona tose); la mejora de la salud de los trabajadores (al detectar problemas posturales o conductas indicativas de problemas en ésta); la prevención de violencia y acoso en el trabajo; la tutela de las trabajadoras víctimas de violencia de género (al identificarse a las personas sobre las que recaigan órdenes de alejamiento); la objetivización de la evaluación del desempeño y del control de calidad; y el control real del tiempo de trabajo, entre otras posibilidades.

En particular se ha señalado su utilidad para asegurar el respeto de la normativa preventiva, detectando posibles incumplimientos en el uso de equipos de protección individual, o en el mantenimiento de las distancias de seguridad. Este uso es especialmente adecuado en el caso de trabajadores que, por la naturaleza de su actividad, prestan sus servicios en solitario. Desde otra perspectiva, resulta muy eficiente para mejorar algunos aspectos de la seguridad para las personas, cuando se aplica a colectivos críticos como conductores o pilotos, detectando problemas físicos o inobservancia de descansos obligatorios y otras restricciones; o alertando de situaciones de violencia física que puedan afectar a usuarios de instalaciones y clientes.

## 5. La videovigilancia en el derecho algorítmico

Una vez identificada la existencia de una combinación entre sistemas de videovigilancia y modelos de IA, que dan lugar a la videovigilancia inteligente y al análisis de la imagen, corresponde señalar, siquiera someramente, cuáles serían las consecuencias jurídicas. Porque, como hemos indicado antes, nos encontramos en una zona de confluencia y superposición de regulaciones, las propias del Derecho Algorítmico con las específicas para esta forma de control elaboradas en el seno del Derecho Digital del Trabajo. Porque las cámaras inteligentes, por el hecho de serlo, no suponen la inaplicación de las reglas que se aplican al conjunto de mecanismos de control de imagen; antes bien, unas y otras se aplicarán de manera simultánea, lo que podría dar lugar a problemas de coordinación entre las distintas regulaciones.

Un dato importante es que el Derecho digital contiene regulaciones específicas para la monitorización por imágenes, lo que no ocurre con el de la IA, que es general y

se ocupa a todo tipo de sistemas. Encontramos algunas reglas específicas para estos dispositivos, pero no una ordenación propia. Esto obliga a un esfuerzo de interpretación y adaptación mayor.

En la práctica, en muchos casos la utilización de cámaras inteligentes va a suponer una acumulación de obligaciones para los empleadores que las usan. También para las empresas que las construyen, comercializan a instalan, puesto que el Reglamento de IA les impone un conjunto elevado de deberes, a diferencia de una regulación de los aspectos digitales que los ignoraba más allá de alguna exigencia de carácter técnico, en normas con este carácter. En esta regulación, y como consecuencia de su naturaleza y objetivos, el foco no se pone sólo en quién utiliza el mecanismo de control sino también en quién se lo proporciona.

Un concepto importante en la aplicación del Derecho Algorítmico de la Unión Europea es el de "espacio de acceso público", que es definido como "*cualquier espacio físico al que pueda acceder un número indeterminado de personas físicas y con independencia de si es de propiedad privada o pública y de la actividad para la que pueda utilizarse el espacio*". No se consideran de esta naturaleza, según el Considerando 19 del Reglamento IA, los locales de empresas y fábricas, así como las oficinas y lugares de trabajo a los que solo se pretende que accedan los empleados y proveedores de servicios pertinentes. Esto condiciona el régimen jurídico de la utilización de los sistemas inteligentes de control de imágenes en numerosos centros de trabajo, aunque habrá otros en los que la presencia de clientes y usuarios dará lugar a esta calificación.

La integración de IA supone que los sistemas de control de imagen deban distinguirse según el riesgo, siguiendo la clasificación en cuatro niveles establecida por el Reglamento. Esto hace que resulten jurídicamente relevantes las finalidades y los usos que se les van a dar a las imágenes que se obtienen, que pueden ser, como se ha visto, múltiples, muchos más que los de control y vigilancia tradicionalmente asociados a este tipo de dispositivos.

El Reglamento de IA incluye entre las prácticas de IA prohibidas algunas que pueden estar entre en las que estamos estudiando. Dentro de los ocho supuestos contemplados por su artículo 5 están los sistemas se usan para inferir las emociones de una persona física en los lugares de trabajo y en los centros educativos (salvo los que estén justificados por motivos médicos o de seguridad). El Reglamento define el concepto de "sistema de reconocimiento de emociones" como aquellos que, utilizando la IA, distinguen o deducen las emociones o las intenciones de las personas físicas a partir de sus datos biométricos. Las imágenes obtenidas bien por circuito cerrado, bien a partir de los ordenadores de los empleados, son una gran fuente de información para ello.

Se excluyen de este concepto los estados físicos como el dolor o el cansancio, y la norma unioneuropea pone como ejemplo los sistemas utilizados para detectar el cansancio de los pilotos o conductores profesionales con el fin de evitar accidentes, lo que permite la analítica de imágenes en el marco de las políticas de seguridad y salud de los trabajadores, seguridad de los usuarios o incluso de bienestar físico general de la plantilla.

Si el sistema de análisis de las imágenes captadas en la empresa va a dar soporte a decisiones de recursos humanos o de producción, o incluso si va a adoptar las mismas decisiones, entonces nos encontraríamos en un supuesto de los contemplados en el Anexo III del Reglamento como de alto riesgo, que como es sabido incluyen los sistemas de IA destinados a ser utilizados para tomar "*decisiones que afecten a las condiciones de las relaciones de índole laboral o a la promoción o rescisión de relaciones contractuales de índole laboral, para la asignación de tareas a partir de comportamientos individuales o rasgos o características personales o para supervisar y evaluar el rendimiento y el comportamiento de las personas en el marco de dichas relaciones*". Idéntica calificación de sistemas de alto riesgo merecen para el Reglamento los de identificación biométrica remota, salvo aquellos cuya finalidad sea exclusivamente de verificación biométrica para confirmar que una persona física concreta es la persona que afirma ser; así como los utilizados para la categorización biométrica y el reconocimiento de emociones.

Tal calificación supone, como es sabido, la exigencia en cascada de una serie de medidas, como la necesidad de implantar un sistema de gestión de riesgos, la elaboración de una documentación técnica, prácticas de gobernanza y gestión de datos adecuadas, un nivel de transparencia suficiente, un registro automático de acontecimientos y una supervisión humana, además de exigencias concretas de precisión, solidez y ciberseguridad.

Va a cambiar radicalmente la naturaleza y el alcance del deber de información a la representación de la plantilla. La LOPDGDD impone en su artículo 89 el deber de la empresa que imponga sistemas de cámaras o videocámaras para el ejercicio de las funciones de control de los trabajadores de informar con carácter previo a éstos y a sus representantes. La inclusión de la IA activa la aplicación del artículo 64.4 TRET, que como es sabido reconoce el derecho de comités de empresa y delegados de personal de ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles. El conocido como "derecho de información algorítmico" existiría respecto de los sistemas de análisis de imágenes vinculados a aparatos para la captación de éstas. Igualmente se aplicaría el artículo 26.7 del Reglamento IA, que impone que antes de utilizar un sistema de IA de alto riesgo en el lugar de trabajo la empresa informe a los representantes de los trabajadores afectados de que estarán expuestos. Esta obligación se cumplirá siguiendo la regulación vigente en el Derecho de la Unión y nacional y conforme a las prácticas en materia de información a los representantes de los trabajadores. Así lo prevé el Considerando 90 del Reglamento, que señala que sus disposiciones son independientes de las que puedan existir en materia de información y consulta a los trabajadores o a sus representantes en virtud del Derecho o las prácticas nacionales.

El deber de información individual a los trabajadores afectados del artículo 89 LOPDGDD se completa con el previsto por el artículo 26.7 del Reglamento IA cuando el control de imágenes se caracteriza como de alto riego por su finalidad y alcance. Para la implementación de este deber la norma unioneuropea se remite igualmente al Derecho

aplicable en materia de información a los trabajadores, y a las prácticas nacionales que pudieran existir. Este derecho de información individual se completa con el derecho a recibir una explicación de decisiones que le afecten previsto en el artículo 86 del Reglamento, lo que supone que el trabajador afectado por una decisión basada en los resultados un sistema que integre IA y captación de imágenes tendrá derecho a obtener del responsable de su implantación, su empleador, explicaciones claras y significativas acerca del papel que ésta ha tenido en todo el proceso decisión y en su resultado.

En relación con la posibilidad de manipulación y generación de imágenes, que puedan hacerse pasar por las obtenidas por el sistema de videovigilancia de la empresa, el artículo 50 del Reglamento de Inteligencia Artificial indica que los proveedores del sistema velarán por que los resultados de salida del sistema de IA estén marcados en un formato legible por máquina y que sea posible detectar que han sido generados o manipulados de manera artificial. Además, éstos deberán velar por que sus soluciones técnicas sean eficaces, interoperables, sólidas y fiables en la medida en que sea técnicamente viable, teniendo en cuenta las particularidades y limitaciones de los diversos tipos de contenido, los costes de aplicación y el estado actual de la técnica generalmente reconocido, según se refleje en las normas técnicas pertinentes.

Finalmente, el uso de sistemas inteligentes de control de imágenes impone a las empresas el deber de alfabetización en materia de IA, en tanto sus empleados van a ser personas afectadas. Ello supone la adquisición de los conceptos necesarios para tomar decisiones con conocimiento de causa en relación con tales sistemas. El Reglamento de IA (Considerando 20) dispone además que la puesta en práctica general de medidas de alfabetización en materia de IA y la introducción de acciones de seguimiento adecuadas podrían contribuir a mejorar las condiciones de trabajo. En general la alfabetización en materia de IA incluye la sensibilización pública y la comprensión de los beneficios, los riesgos, las salvaguardias, los derechos y las obligaciones en relación con el uso de sistemas de IA.

## 6. Reflexiones conclusivas

La IA se está desarrollando y se está extendiendo por todos los entornos de trabajo, afectando a la forma en que se prestan los servicios y se organiza la actividad productiva. También al modo en que se supervisa a las personas que trabajan. Esto se hace, en muchos casos, integrando los dispositivos tradicionalmente utilizados en esta tarea con programas que optimizan su uso y manejan la información suministrada por éstos. La videovigilancia se convierte, así, en un control inteligente, que hace posible además una analítica de imágenes.

Este avance técnico incrementa los riesgos para los trabajadores, ya importantes en un control de por sí intrusivo. También mejora sus utilidades, que incluyen algunas que debemos considerar positivas para los trabajadores y la sociedad en su conjunto, en aspectos tales como la salud y seguridad laborales y la seguridad de las personas. Esto

obliga a un análisis del marco normativo de aplicación, que es lo que se ha pretendido hacer en estas páginas.

Pero los efectos de los avances algorítmicos en el control de imágenes van más allá, ya que se convierten en un instrumento adicional en el proceso de toma de decisiones previo al ejercicio del poder de dirección. De la misma manera en que, como dijimos al principio de estas páginas, la revolución tecnológica está haciendo cambiar el peso real de los dos principales poderes empresariales en las relaciones laborales, el de dirección y el de control, potenciando la supervisión de personas que prestan sus servicios con mayor autonomía, la utilidad de estos mecanismos está mutando. Junto a la original de velar por el cumplimiento de las obligaciones laborales de los empleadores, las nuevas utilidades están permitiendo que sean instrumentales también para el ejercicio del poder de dirección. Y ello porque dan soporte a decisiones empresariales de todo tipo con la mejora de la información obtenida, y con el tratamiento de ésta para identificar patrones y pautas, predecir acontecimientos y proponer actuaciones.

La videovigilancia inteligente es útil también para una función de creciente importancia, la del control del cumplimiento de obligaciones de empresarios y trabajadores, porque puede orientarse a la identificación de posibles violaciones de éstas. En momentos en los que el compliance adquiere una mayor relevancia en las empresas, este apoyo merece ser destacado.

Estos mecanismos suponen una zona de confluencia entre el Derecho Digital tradicional y el Derecho Algorítmico, en la medida en que ambos sectores del ordenamiento resultan aplicables. La acumulación de obligaciones que se deriva de ello no generará, a mi juicio, grandes problemas, en la medida en que resultan compatibles. Unos mismos instrumentos aparecen en las dos regulaciones, aunque la de la IA va más allá y contiene otros de nueva creación. El origen común y el hecho de compartir unas mismas finalidades de tutela de las personas explica esta compatibilidad. Sí será necesario, por supuesto, una mayor coordinación. Seguramente también más avances en la regulación algorítmica, que como se dijo en su momento no ha producido hasta ahora reglas específicas para el mecanismo de control que nos ocupa; es de esperar que comencemos a verlas a medida que la utilización de estos productos se vaya extendiendo.

Estos avances resultan especialmente recomendables, a mi juicio, porque la videovigilancia inteligente debe disponer un marco regulatorio adecuado que le permita extenderse y alcanzar todas sus posibilidades, evitando a la vez los riesgos para las personas que trabajan sometidos a ella. Estamos en un momento de transición tecnológica, que está siendo acompañado de una reordenación normativa que, en esta materia como en otras, debe esforzarse para mantenerse a la altura.

# Los sistemas automatizados de reconocimiento de emociones en el trabajo en el reglamento europeo de inteligencia artificial

## Automated emotion recognition systems at work in the European Artificial Intelligence Act

Ana Belén Muñoz Ruiz

*Profesora Titular de Derecho del Trabajo y de la Seguridad Social (catedrática acreditada)*
*Universidad Carlos III de Madrid*

ORCID ID: 0000-0002-8863-9938

doi: 10.20318/labos.2024.9033

*Resumen:* Desde los comienzos de los años 90 se viene investigando en los sistemas de reconocimiento automatizados de las emociones. Los sistemas automatizados de reconocimiento de emociones basados en los datos biométricos (rostro, voz, movimiento corporal, entre otros) pueden ser una fuente potencial de información para las empresas sobre sus trabajadores. En el artículo se estudia la nueva regulación de estos sistemas en el Reglamento europeo de Inteligencia Artificial con un doble propósito: primero, identificar el alcance y contenido de las obligaciones y prohibiciones de las empresas que1 tratan esta tipología de datos de carácter personal con apoyo en la inteligencia artificial; y segundo, comprender que los sistemas de reconocimiento de emociones en el ámbito laboral suponen un salto cualitativo si los comparamos con los controles tradicionales (videovigilancia, email, Internet, ordenador y dispositivos semejantes). Como se verá, este tipo de control de nueva generación sitúa a las personas trabajadoras en una vulnerabilidad extrema y precisan de mayores garantías.

*Palabras clave:* Reglamento europeo, biometría, dato de carácter personal, sistema automatizado de reconocimiento de emociones, inteligencia artificial, derecho del trabajo, obligaciones empresariales, intimidad.

*Abstract:* Since the early 1990s, research has been carried out on automated emotion recognition systems. These automated emotion recognition systems, which are based on biometric data (face, voice, body movement, etc.), may be a potential source of worker information for companies. The paper analyses the new European regulation which includes those mentioned systems on twofold purpose. First, to identify the scope and content of the obligations and prohibitions of companies that process this type of personal data with the support of artificial intelligence; and second, to explain that emotion recognition systems in the labor setting are a qualitative leap as compared to traditional control measures (video surveillance, email, Internet, computer and similar devices). It is shown that these new generation control measures may place employees in situations of extreme vulnerability and therefore require additional guarantees.

*Keywords:* European Artificial Intelligence Act, biometrics, personal data protection, automated emotion recognition systems, artificial intelligence, labour law, duties of the employer, privacy.

## 1. Introducción

Desde los comienzos de los años 90 se viene investigando en los sistemas de reconocimiento automático de las emociones. Programadores y matemáticos sugieren tres modos de expresión emocional adecuados para la detección automatizada: (i) la emoción a partir de la expresión facial; (ii) la detección de la emoción a partir del habla y; (iii) la detección de la emoción multimodal, es decir, la combinación de la emoción facial y del habla. De los tres modos de expresión emocional, la expresión facial es una de las formas más poderosas que tienen las personas para entablar conversaciones y comunicar emociones y otras señales mentales, sociales y fisiológicas. Una de las formas más importantes en que las personas muestran sus emociones es a través de las expresiones faciales[1].

Cuando nos referimos a emociones comúnmente nos referimos a emociones concretas y a los estados de ánimo (por ejemplo, estados depresivos o ansiosos). ¿Y los sentimientos? Los sentimientos se definen como el poso que dejan las emociones[2].

Querer comprender el comportamiento humano ignorando las emociones es como querer comprender el funcionamiento de un coche ignorando su motor. Los sistemas de reconocimiento de emociones pueden ser una fuente potencial de información para las empresas sobre las personas trabajadoras[3]. En esta lógica, resulta igual de importante comprender las propias emociones como reconocer las emociones de los demás. Más aún cuando, con frecuencia, las motivaciones humanas residen en el estado emocional de los agentes[4].

Precisamente el nuevo Reglamento (UE) 2024/1689, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (en adelante, RIA), introduce el concepto de reconocimiento automatizado de emociones para establecer unos mecanismos de tutela en el ámbito laboral. Si bien se establece una entrada en vigor aplazada y distinta en función del articulado de esta norma[5], la novedad de su regulación es motivo suficiente para realizar un análisis pormenorizado del concepto legal de sistema automatizado de emociones, sus implicaciones para los trabajadores y los niveles de riesgo que introduce.

## 2. El concepto legal de sistema automatizado de reconocimiento de emociones y los motivos para su inclusión

Como se ha anticipado, el RIA da un paso adelante e introduce por primera vez en una norma jurídica de alcance comunitario el concepto de sistema automatizado de recono-

[1] K. Prasanthi Jasmine y K. Naga Prakash, *Reconocimiento de emociones humanas a partir de imágenes de rostros*, Ediciones Nuestro Conocimiento, 2021, pp. 5-6.

[2] V. Camps, *El gobierno de las emociones*, Herder, Barcelona, 2011, p. 40.

[3] Sobre este tema, me remito a mi libro A.B., Muñoz Ruiz, *Biometría y sistemas automatizados de reconocimiento de emociones: implicaciones jurídico-laborales*, Tirant Lo Blanch, Valencia, 2023.

[4] D. Pinea Oliva, *Sobre las emociones*, Ediciones Cátedra, 2019, p. 12.

[5] Con carácter general la entrada en vigor está prevista para el 2 de agosto de 2026. Vid. artículo 113 del RIA.

cimiento de emociones. Se define el sistema de reconocimiento de emociones como un sistema de inteligencia artificial (en adelante, IA) destinado a distinguir o inferir las emociones o las intenciones de las personas físicas a partir de sus datos biométricos (artículo 3 (39) RIA)[6]. Sustituir el texto tachado por este otro: En la condición de responsables del despliegue se enumeran en el RIA las obligaciones de las empresas y entidades públicas que usen estos sistemas de alto riesgo y que son las siguientes: a) Deber de transparencia y explicación individual. Transparencia informando al trabajador afectado y a la representación de los trabajadores de que están expuestos a este tipo de sistema. Dicha información debe proporcionarse con anterioridad a la puesta en servicio o utilización del sistema de IA en el lugar de trabajo (artículo 26.7 RIA). Por su parte, la explicación individual consiste en el derecho del trabajador (y obligación de la empresa) a recibir 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 cuando 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 (artículo 86 RIA). b) Cuando proceda, los responsables del despliegue de sistemas de IA de alto riesgo utilizarán la información facilitada conforme al artículo 13 del RIA para cumplir la obligación de llevar a cabo una evaluación de impacto relativa a la protección de datos que les imponen el artículo 35 del RGPD o el artículo 27 de la Directiva (UE) 2016/680, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos. c) Supervisión humana. Los sistemas de IA de alto riesgo precisan de supervisión humana y los responsables del despliegue deben encomendar dicha supervisión a las personas físicas que tengan la competencia, la formación y la autoridad necesarias (artículo 26.2 RIA). d) Deber de vigilar el correcto funcionamiento de los sistemas sobre la base de las instrucciones de uso y, cuando proceda, informarán al proveedor o distribuidor y a la autoridad competente de acuerdo con el artículo 72 del Reglamento,

[6] En la tramitación del Reglamento comunitario se formularon algunas propuestas de cambio respecto del concepto originario que, finalmente, no prosperaron. La versión originaria definía el sistema automatizado de emociones del siguiente modo: "el sistema de reconocimiento de emociones es un sistema de inteligencia artificial (en adelante, IA) destinado a detectar o deducir las emociones o las intenciones de personas físicas a partir de sus datos biométricos". El 25.11.2022 el Consejo de la Unión Europea adoptó su posición (también denominada "orientación general") sobre la Ley de Inteligencia Artificial. En dicho documento se realizan algunas modificaciones en la definición de sistema de reconocimiento de emociones. Se propuso la siguiente definición: "un sistema de IA destinado a detectar o deducir los estados mentales, las emociones o las intenciones de las personas físicas a partir de sus datos biométricos". Como se observa, se incluyen los estados mentales para así dar cobertura a los estados de ánimo que podrían no considerarse emociones, por ejemplo, estar confuso, estar despistado, falta de concentración. Además de estados mentales, se podría añadir estados físicos (por ejemplo, estar cansado, tener una cojera, trabajador con una lesión). Con fecha de 14 de junio de 2023 el Parlamento Europeo incorporó en la definición los pensamientos: "el sistema de reconocimiento de emociones es un sistema de IA destinado a detectar o deducir las emociones, los pensamientos, los estados de ánimo o las intenciones de individuos o grupos a partir de sus datos biométricos y sus datos de base biométrica".

relativo al sistema de vigilancia poscomercialización. e) Los responsables del despliegue que sean autoridades públicas o instituciones, órganos y organismos de la Unión deben cumplir las obligaciones de registro previstas en el artículo 49 del Reglamento.

La definición de dato biométrico a que se refiere el RIA coincide con la establecida en el artículo 4 (14) del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo 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), según el cual los datos biométricos pertenecen a la categoría de datos especiales que se obtienen partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos. Otros ejemplos de datos biométricos son el reconocimiento de iris o retina, firma, voz, etc.[7]

En el RIA el concepto de emociones 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. Pero no incluye los estados físicos, como el dolor o el cansancio, como, por ejemplo, los sistemas utilizados para detectar el cansancio de los pilotos o conductores profesionales con el fin de evitar accidentes. Tampoco incluye la mera detección de expresiones, gestos o movimientos que resulten obvios, salvo que se utilicen para distinguir o deducir emociones. Esas expresiones pueden ser expresiones faciales básicas, como un ceño fruncido o una sonrisa; gestos como el movimiento de las manos, los brazos o la cabeza, o características de la voz de una persona, como una voz alzada o un susurro (Considerando 18 del RIA).

En definitiva, la emoción de una persona expone su vulnerabilidad esencial. Se dice que un ser sin emociones porque se ha liberado de éstas no es un ser humano[8]. De hecho, los ordenadores emocionales que simulan intencionalidad, emociones, valores y sentido común son eso, únicamente simulaciones, no realidades. Hacen como si sintieran, pero para sentir se necesita un cuerpo[9].

¿Por qué abordar los sistemas de reconocimiento de emociones en el RIA? La razón se explica en el preámbulo de la norma cuando se dice que los datos biométricos pueden permitir las funciones tradicionales (autenticación, la identificación o la categorización de las personas físicas) pero también el reconocimiento de las emociones de las personas físicas (Considerando 14 RIA). Y esto es preocupante porque el tratamiento de datos biométricos por parte de las empresas puede derivar en el conocimiento de enfermedades de la persona trabajadora o predisposición a padecerlas sin cumplir la finalidad de prevención de riesgos laborales. A su vez, pueden producirse fallos del sistema de la IA debido a las singularidades culturales y derivar en discriminaciones[10].

[7] Vid. Artículo 3. 34) RIA.

[8] V. Camps, *El gobierno de las emociones*, Herder, Barcelona, 2011, p. 38.

[9] A. Cortina, Orts, "Ética de la inteligencia artificial", *Anales de la Real Academia de Ciencias Morales y Políticas*, 2019, nº 96, p. 385.

[10] AB, Muñoz Ruiz, *Biometría y los sistemas automatizados de reconocimiento de emociones: implicaciones jurídico-laborales*, Tirant Lo Blanch, Valencia, 2023.

A diferencia de los controles tradicionales (videovigilancia, ordenador, Internet, entre otros), los sistemas de reconocimiento de emociones emplean algoritmos e IA que, como se verá, incrementará la capacidad de análisis y explotación del resultado alcanzado. Lo que va a significar una mayor intromisión en los derechos de intimidad y protección de datos de carácter personal de las personas trabajadoras y producir lesiones indirectas de otros derechos fundamentales (discriminación, daños a la salud mental y su conexión con la seguridad y salud en el trabajo).

Los datos biométricos (voz, rostro, movimiento corporal, entre otros) han experimentado una profunda transformación. Por lo que se refiere a la voz, los científicos han conseguido con varias técnicas de procesado capturar esta capa de información, oculta a primera vista, amplificando y extrayendo características tonales y acústicas del habla humana. Las emociones que este sistema puede detectar pueden ser calma, felicidad, tristeza, ira, temor, asco, sorpresa… o simplemente "neutral"[11]. En este sentido, se describen supuestos donde las máquinas juzgan a los humanos y, por ejemplo, pueden tener como resultado que una persona no consiga un trabajo por el tono de voz[12].

La fiabilidad de este sistema se ve comprometida por la presencia de "Ruido" en la grabación, a pesar de ser un procedimiento computacionalmente más sencillo que el reconocimiento del habla ("Ok Google", Alexa, Siri….)[13]. De todas formas, los valores de fiabilidad esperados en los resultados superan el 80% y pueden llegar a alcanzar el 95%, dependiendo de la calidad de la fuente de grabación y de su procesamiento.

El análisis de la voz por sí mismo no es capaz de analizar ningún aspecto de la salud ("física") de la persona propietaria de la voz, incluso la salud "psíquica" es de muy difícil interpretación usando sólo la voz, pero eso no quiere decir que no pueda ser de ayuda en la salud de las personas. Por ejemplo, en Taiwán se realizó un ensayo en el cual se utiliza este sistema para el análisis de las reacciones de los pacientes durante las consultas. Esto sirve para "entrenar" a los doctores a tener una mayor empatía. El objetivo es conseguir este análisis y un asesoramiento al facultativo a tiempo real[14].

Existe consenso científico en relación a que al afirmar que es el rostro donde están ubicados muchos de los rasgos en que los humanos nos fijamos para reconocer a otro, juega además un papel clave en la comunicación e interacción con los demás, en la transmisión de la identidad y de la emoción[15]. Precisamente, el reconocimiento facial de emociones (FER/Facial Emotion Recognition) es la tecnología que analiza las expresio-

[11] https://www.projectpro.io/article/speech-emotion-recognition-project-using-machine-learning/573

[12] Vid. F. Pasquale, *Las nuevas leyes de la robótica. Defender la experiencia humana en la era de la IA*, Galaxia Gutenberg, S.L. 2024, p. 173.

[13] Instituto Nacional de Ciberseguridad (INCIBE), Tecnologías biométricas aplicadas a la ciberseguridad. Una guía de aproximación para el empresario, 2016, p. 12; accesible en: https://www.incibe.es/sites/default/files/contenidos/guias/doc/guia_tecnologias_biometricas_aplicadas_ciberseguridad_metad.pdf.

[14] https://www.mdpi.com/2076-3417/11/11/4782

[15] L, Escajedo San Epifanio, *Reconocimiento e identificación de las personas mediante Biometrías estáticas y dinámicas*, Tesis Doctoral, Universidad de Alicante, 2015, p. 86.

nes faciales, tanto de imágenes como de vídeos, con el objetivo de obtener información acerca del estado emocional del sujeto[16].

Debido a que la cara es uno de nuestros medios más importantes de comunicación no verbal, la cantidad de emociones detectables por este medio es alta: ira, asco, miedo, alegría, sorpresa… Los modelos pueden incluso detectar "emociones compuestas", por ejemplo "tristemente sorprendido", "sorprendentemente enfadado"….

Se sabe que la comunicación verbal supone solamente un 7% de la comunicación humana, mientras que la expresión facial supone entre un 38% y un 55%[17].

En cuanto al entorno laboral, esta técnica tendrá uso en dos vertientes: aspectos puramente laborales y detección de enfermedades de la persona trabajadora. En cuanto a los aspectos puramente laborales, esta tecnología ayudará a los reclutadores después de las entrevistas de trabajo a tomar decisiones (y en un futuro durante las entrevistas), detectará el interés del candidato durante la entrevista de trabajo y durante el trabajo, se podrá monitorizar las actitudes, atención y motivación de los empleados.

Los datos de salud que se pueden detectar con esta técnica son limitados, lógicamente se limitan a enfermedades mentales: detección de autismo, enfermedades degenerativas, trastornos psicóticos, tendencias suicidas, depresión…. También está en estudio la posible detección de enfermedades genéticas que tengan un reflejo en la cara de la persona.

Por lo que se refiere al estudio de emociones a través de los movimientos del trabajador, en el ámbito laboral es posible la detección de estados de ansiedad y de estrés[18]. El gran problema de desarrollo de esta técnica es la identificación de las características relacionadas con las emociones en los movimientos del cuerpo humano, es decir, relacionar ciertos movimientos o conjuntos de movimientos con emociones, o un conjunto de emociones. Las emociones que se pueden estudiar aquí son, de momento, felicidad, tristeza, miedo, ira y "neutral". La gran ventaja de este sistema es que el sujeto objeto de estudio puede ni siquiera saber que lo está siendo, ya que la imagen puede capturarse a gran distancia (no como el análisis de la voz, o de la expresión facial).

La precisión de estos sistemas llega a valores del 90%, incluso del 96% si la persona está sentada (y el algoritmo está preparado para ello). Si la persona puede estar haciendo diferentes acciones, la precisión del algoritmo puede bajar, teniendo un valor aproximado del 85%. De todas formas, el sistema tiene dificultades que todavía no han sido corregidas, por ejemplo, si la persona está andando[19].

---

[16] Vid. sobre el tema: https://edps.europa.eu/system/files/2021-05/21-05-26_techdispatch-facial-emotion-recognition_ref_en.pdf

[17] C. Blushan y otros, Facial Expression Recognition, https://www.wsj.com/articles/BL-DGB-42522

[18] https://www.researchgate.net/publication/338238356_Emotion_Recognition_From_Body_Movement

[19] Como ejemplo, se ha publicado noticia de un proyecto piloto que se desarrollará en las cárceles de Cataluña. La prueba piloto consiste en analizar a través de las imágenes registradas por las cámaras de vigilancia interna y de la IA las expresiones faciales y el lenguaje corporal de los reclusos. El objetivo es prevenir riesgos que puedan producirse, como una fuga o la introducción de droga en la prisión. La implantación de este sistema ha sido adjudicada a una empresa, que desarrolla proyectos detección biométrica. La empresa adjudicataria creará e instalará un sistema automatizado de identificación facial y control de movimientos de internos en zonas críticas del perímetro de seguridad del centro. El sistema servirá para realizar búsqueda de datos y su posterior

## 3. Los niveles de riesgo en el nuevo Reglamento europeo de Inteligencia Artificial

En el articulado del RIA se palpa el carácter inspirador de la propuesta legislativa alemana. La Comisión Ética de Datos constituida por el Gobierno alemán recomendó en 2019 adoptar un enfoque normativo basado en el riesgo distinguiendo cinco niveles de criticidad en función de las variables de probabilidad y severidad del daño como consecuencia del empleo de algoritmos. A diferencia de algunos informes que ponen el foco en las regulaciones nacionales, el documento de trabajo mencionado afirma que se precisa una nueva regulación europea sobre sistemas algorítmicos fijando unos requisitos generales horizontales que deberían ser desarrollados por normas sectoriales (entre ellas, el derecho del trabajo) [20].

Siguiendo este enfoque de regulación basado en el riesgo, el RIA menciona de forma expresa que el límite de los sistemas de IA son los derechos fundamentales: dignidad, intimidad, protección de datos de carácter personal, no discriminación y seguridad y salud de los ciudadanos y también de las personas trabajadoras.

En efecto, el RIA, a diferencia de otras normas previas con una lógica semejante como el RGPD, donde las referencias a la salud se concentran en el tratamiento de los datos de salud al ser considerado como datos especialmente protegidos, establece como uno de sus objetivos garantizar un nivel elevado de protección de la salud y seguridad: "Artículo 1º: El objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior y 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 consagrados en la Carta, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA (en lo sucesivo, «sistemas de IA») en la Unión así como prestar apoyo a la innovación".

Se afirma en el Considerando (47) del RIA que "los sistemas de IA pueden tener un efecto adverso para la salud y la seguridad de las personas, en particular cuando funcionan como componentes de seguridad de productos. (…) Por ejemplo, los robots cada vez más autónomos que se utilizan en las fábricas o con fines de asistencia y atención personal deben poder funcionar y desempeñar sus funciones de manera segura en entornos complejos. Del mismo modo, en el sector sanitario, donde puede haber repercusiones especialmente importantes en la vida y la salud, los sistemas de diagnóstico y de apoyo a las decisiones humanas, cuya sofisticación es cada vez mayor, deben ser fiables y precisos".

---

análisis para clasificar los perfiles que presenten riesgo de violencia en el interior de los recintos penitenciarios. A través de esta tecnología también se podrá evaluar si algún interno introduce droga u objetos prohibidos en la cárcel. El reconocimiento gestual también permitirá analizar expresiones, actitudes o comportamientos de los presos, incluso tras una comunicación íntima o con la familia (una vis a vis). En la actualidad, este control dependía, en gran parte, del conocimiento que tienen los funcionarios. Vid. noticia: https://www.elperiodico.com/es/sociedad/20230920/carceles-cataluna-inteligencia-artificial-control-presos-92321451

[20] Más extensamente sobre la propuesta alemana me remito a mi trabajo A.B., Muñoz Ruiz, ¿Se deben regular los algoritmos? Un breve análisis a la propuesta normativa alemana: la pirámide de criticidad basada en el riesgo, *Blog El Foro de Labos*, 11.12.2019.

Algunos estudios advierten que la tecnología de control emocional puede tener implicaciones para la salud y seguridad de las personas trabajadoras. Esta tecnología permite conocer los movimientos corporales, los signos vitales, los indicadores de estrés y fatiga, las micro expresiones faciales, el tono de voz y análisis de sentimiento. De hecho, pueden dar lugar a que los trabajadores pierdan el control de sus puestos de trabajo y aumenten la micro gestión, la presión por el rendimiento, la competitividad, la individualización y el aislamiento social. Al sentir que los trabajadores que su privacidad está siendo invadida, puede generar ansiedad y estrés. Es probable que los propios trabajadores declinen tomar descansos cuando los necesitan, lo que puede causar accidentes y problemas de salud, como trastornos musculoesqueléticos y enfermedades cardiovasculares. Los horarios de trabajo inestables, como los horarios a corto plazo establecidos automáticamente por algoritmos, pueden tener impacto negativo en los trabajadores, incluido un mayor conflicto entre el trabajo y la familia y el estrés laboral y la incertidumbre de ingresos [21].

El RIA supone un salto cualitativo respecto del RGPD por varias razones. En primer lugar, se da el paso de aprobar un marco jurídico de la IA que supone una mayor envergadura que el procesamiento de datos de carácter personal. En segundo lugar, se concede un especial protagonismo a los datos biométricos (rostro, voz, huella dactilar, movimientos corporales, entre otros) y sus nuevas formas de captación (remota, en tiempo real, en tiempo diferido, entre otros). Y tercera, se supera el plano de los datos de carácter personal y se incluyen las emociones de las personas trabajadoras.

Con la aprobación del RIA se incrementa la protección de los datos de salud de las personas trabajadoras. Es cierto que no se modifican de forma explícita las reglas y garantías de tratamiento de la salud de las personas empleadas previstas en el RGPD y en la Ley 31/1995, de 8 noviembre, de Prevención de Riesgos Laborales (LPRL). Sin embargo, se da un paso más en el RIA al incluir protección frente a los sistemas automatizados de reconocimiento de emociones de las personas.

A partir de estas premisas, el RIA establece cuatro niveles de riesgo donde podemos encontrar alguna expresión de los sistemas automatizados de reconocimiento de emociones de las personas empleadas.

En primer lugar, los sistemas de reconocimiento de emociones aparecen mencionados de forma explícita en el nivel de riesgo 1 (riesgo inaceptable), luego, están prohibidos. Sin embargo, se recogen dos excepciones. Señala el artículo 5.1 f) RIA: "Quedan prohibidas las siguientes prácticas de IA: (…) La introducción en el mercado, la puesta en servicio para este fin específico o el uso de sistemas de IA para inferir las emociones de una persona física en los lugares de trabajo y en los centros educativos, excepto cuando el sistema de IA esté destinado a ser instalado o introducido en el mercado por motivos médicos o de seguridad".

Se siguen de esta forma las directrices aportadas por el Comité Europeo de Protección de Datos (CEPD) y el Supervisor Europeo de Protección de Datos (SEPD) que

[21] OSHA-EU, Impact of artificial intelligence on occupational safety and health, 2021.

habían considerado que el uso de la IA para inferir emociones de una persona física es muy indeseable y deberá prohibirse, excepto en determinados casos de uso bien especificados, a saber, con fines de salud o investigación (por ejemplo, pacientes para quienes el reconocimiento emocional es importante), siempre con las salvaguardias adecuadas y, por supuesto, con sujeción a todas las demás condiciones y límites de protección de datos, incluida la limitación de la finalidad[22].

En segundo término, se debe añadir que los sistemas de reconocimiento de emociones también podrían estar en el nivel de riesgo 2: sistemas de IA de alto riesgo cuando no estén prohibidos, por ejemplo, aquéllos que estén justificados por motivos médicos o de seguridad. De hecho, en el Anexo III del RIA donde se enumeran los sistemas de IA de alto riesgo aparecen los sistemas de IA destinados a ser utilizados para el reconocimiento de emociones y se dice en el Considerando (54) del RIA que: "Además, deben clasificarse como de alto riesgo los sistemas de IA destinados a ser utilizados para la categorización biométrica conforme a atributos o características sensibles protegidos en virtud del artículo 9, apartado 1, del Reglamento (UE) 2016/679 sobre la base de datos biométricos, en la medida en que no estén prohibidos en virtud del presente Reglamento, así como los sistemas de reconocimiento de emociones que no estén prohibidos con arreglo al presente Reglamento".

En la condición de responsables del despliegue se enumeran en el RIA las obligaciones de las empresas y entidades públicas que usen estos sistemas de alto riesgo y que son las siguientes: a) Deber de transparencia y explicación individual. Transparencia informando al trabajador afectado y a la representación de los trabajadores de que están expuestos a este tipo de sistema. Dicha información debe proporcionarse con anterioridad a la puesta en servicio o utilización del sistema de IA en el lugar de trabajo (artículo 26.7 RIA). Por su parte, la explicación individual consiste en el derecho del trabajador (y obligación de la empresa) a recibir 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 cuando 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 (artículo 86 RIA). b) Cuando proceda, los responsables del despliegue de sistemas de IA de alto riesgo utilizarán la información facilitada conforme al artículo 13 del RIA para cumplir la obligación de llevar a cabo una evaluación de impacto relativa a la protección de datos que les imponen el artículo 35 del RGPD o el artículo 27 de la Directiva (UE) 2016/680, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos. c) Supervisión humana. Los sistemas de IA de alto riesgo precisan de supervisión humana y los responsables del despliegue deben encomendar dicha supervisión a las personas físicas que tengan la competencia, la formación y la autoridad nece-

[22] CEPD-SEPD Dictamen conjunto 5/2021 sobre Ley de Inteligencia Artificial, 18.6.2021.

sarias (artículo 26.2 RIA). d) Deber de vigilar el correcto funcionamiento de los sistemas sobre la base de las instrucciones de uso y, cuando proceda, informarán al proveedor o distribuidor y a la autoridad competente de acuerdo con el artículo 72 del Reglamento, relativo al sistema de vigilancia poscomercialización. e) Los responsables del despliegue que sean autoridades públicas o instituciones, órganos y organismos de la Unión deben cumplir las obligaciones de registro previstas en el artículo 49 del Reglamento.

Si bien pensamos que los ejemplos que podrían encajar en estas excepciones son limitados (en la medida que el RIA ha excluido los sistemas utilizados para detectar el cansancio de los pilotos o conductores profesionales de la definición de sistema de reconocimiento de emociones, podemos mencionar alguno. Sería el caso antes mencionado de las cámaras de IA que algunas cárceles españolas han implantado para identificar expresiones faciales y lenguaje corporal de los reclusos[23].

En tercer lugar, también podríamos entender que habrá sistemas de este tipo que encajen en el nivel de riesgo 3 (nivel de riesgo limitado) cuando no se basen en los datos biométricos de las personas empleadas. El artículo 50 del RIA se refiere a la IA que se destina a interactuar con personas físicas tales como los robots de software (por ejemplo, *chatbots*). Desde nuestra perspectiva, si el robot de software se apoya en el lenguaje escrito para inferir emociones o intenciones estaríamos en el nivel de riesgo 3 y se aplicarán las obligaciones del artículo 50 RIA (deber de transparencia). Según el artículo 50 RIA los proveedores garantizarán que los sistemas de IA destinados a interactuar directamente con personas físicas se diseñen y desarrollen de forma que las personas físicas de que se trate estén informadas de que están interactuando con un sistema de IA, excepto cuando resulte evidente desde el punto de vista de una persona física razonablemente informada, atenta y perspicaz, teniendo en cuenta las circunstancias y el contexto de utilización.

Y el artículo 50.3 RIA indica que los responsables del despliegue de un sistema de reconocimiento de emociones o de un sistema de categorización biométrica informarán del funcionamiento del sistema a las personas físicas expuestas a él y tratarán sus datos personales de conformidad con los Reglamentos (UE) 2016/679 y (UE) 2018/1725 y con la Directiva (UE) 2016/680, según corresponda.

Ahora bien, si en la conversación entre el chatbot y la persona trabajadora procesa la voz de la persona física (dato biométrico) podríamos estar en el nivel de riesgo 1 o 2 según las circunstancias.

En efecto, existe un tipo de robótica denominada robots de software que trabaja en las sombras y que puede afectar a los derechos de privacidad de las personas empleadas. Nos referimos a los robots RPA (Automatización robótica de procesos) y a los chatbots que habitan en nuestros ordenadores y smartphones y sobre todo tienen la capacidad de hablar, escuchar, reconocernos y contestarnos.

Se han definido como los trabajadores virtuales de cuello azul frente a la robótica industrial que se asimila más a los trabajadores de cuello blanco. A diferencia de los ro-

[23] https://www.elperiodico.com/es/sociedad/20230920/carceles-cataluna-inteligencia-artificial-control-presos-92321451

bots industriales y los coches robóticos, los robots software no son directamente visibles y no tienen una realidad física (en el sentido que no pesan, no ocupan espacio), sino que son pura lógica, puro software. Si se prefiere decirlo de otra manera, son solo programas. Por eso ni los vemos ni les podemos tocar. Se dice que son habitantes de las sombras y que actúan sobre las aplicaciones o los ficheros[24].

Podemos distinguir dos clases de robots software. Por un lado, los RPA o lo que es lo mismo la automatización robótica de procesos. Los robots RPA se dedican a trabajar con otros activos digitales. Fundamentalmente, interactúan con las pantallas de otras aplicaciones y con documentos ofimáticos como hojas de cálculo o ficheros PDF. Lo que hacen, dicho de forma simplificada, es leer datos de pantallas y de ficheros, realizar cálculos o tomar decisiones basadas en esa información y volver a escribir en pantallas o ficheros los resultados o conclusiones obtenidas[25].

De otra parte, existe una categoría de robots que normalmente se denominan chatbots. Un chatbot es un módulo software cuya misión es interaccionar con personas de forma abierta y natural mediante conversaciones. Se trata de una variedad de robots especializados en dialogar con personas mediante conversaciones naturales. Es decir, en esta categoría de robots incluimos aquellos chatbots que interaccionan con nosotros a través de sistemas de mensajería como Facebook Messenger, Slack o WhatsApp, pero también aquellos otros que hablan con nosotros, emiten voz y escuchan y entienden, a su vez, la voz humana. Es decir, incluimos también a los a veces denominados voicebots y los altavoces inteligentes. Nos estamos refiriendo con ello a software del tipo de Alexa, Siri, Cortana y todo tipo de robots más especializados construidos con capacidad de mantener conversaciones por voz [26].

Tanto los RPA como los chatbots incluyen elementos de inteligencia artificial para reconocimiento de voz, para procesamiento de lenguaje natural, para visión artificial y para reconocimiento óptico de caracteres. Por tanto, estos robots, aparte de la programación, incluyen dosis mayores o menores de adaptación a través de lo que denominamos como aprendizaje[27].

En todo caso, los robots de software pueden leer el correo electrónico y los documentos que se procesan en el puesto de trabajo pudiendo llegar a pulverizar el derecho fundamental de intimidad de la persona trabajadora.

Por último, el nivel de riesgo 4 (riesgo mínimo) incluye entre otros videojuegos con IA o filtros de *spam*. En este nivel de riesgo podríamos mencionar la aventura

[24] La doctrina especializada ha identificado seis características de los robots que se aplican a los robots industriales, biológicos y robots software. Se refieren a la artificialidad, capacidad de adaptación, interacción con el entorno, autonomía, sustitución de personas y similitud con la forma de trabajar las personas. Vid. I. G.R., Gavilán, *Robots en la sombra. RPA, robots conversacionales y otras formas de automatización cognitiv*a, Anaya, Madrid, 2021.

[25] Más extensamente G.R., Gavilán, *Robots en la sombra. RPA, robots conversacionales y otras formas de automatización cognitiv*a, Madrid, Anaya, 2021.

[26] I.G.R., Gavilán, *Robots en la sombra. RPA, robots conversacionales y otras formas de automatización cognitiva*, Anaya, Madrid, 2021, pp. 55 y 88.

[27] I.G.R., Gavilán, *Robots en la sombra. RPA, robots conversacionales y otras formas de automatización cognitiva*, Anaya, Madrid, 2021, pp. 55-56.

gráfica y el uso de la gamificación con finalidad laboral. Este tipo de tecnología puede provocar en las personas empleadas distintas emociones durante el proceso como alegría, frustración, decepción, triunfo…

## 4. El impacto del nuevo Reglamento europeo de Inteligencia Artificial en España

El impacto del nuevo RIA en España es notable por dos razones principales. En primer lugar, si bien no se suscitado hasta el momento casuística en nuestro país sobre los sistemas automatizados de emociones, algunas empresas del sector del telemarketing están desarrollando programas piloto de control laboral y calidad con apoyo en la IA. Es decir, primero se graban las llamadas entre los clientes y las personas trabajadoras. A continuación, la IA evalúa estas llamadas conforme a los parámetros fijados por la empresa y, finalmente, un trabajador se encarga de verificar si la IA ha cometido errores.

En Europa se han identificado supuestos de reconocimiento automatizado de emociones cuyo interés adquiere más relevancia tras la aprobación del RIA. En el momento de los hechos no estaba aprobado el RIA y se aplicó el RGPD. En efecto, uno de los primeros supuestos aborda el dato biométrico de la voz y fue resuelto por la resolución de la Agencia Húngara de Protección de Datos de 8 febrero 2022. Los hechos describen a un banco que utilizó un software de procesamiento de señales de voz basado en inteligencia artificial. El período de datos del procesamiento fue de 45 días con respecto a la grabación de sonido que se puede escuchar dentro del software y un año con respecto a las estadísticas y listas de llamadas clasificadas generadas a través de la operación de software. El software analizaba y evaluaba los estados emocionales de los clientes y de las personas empleadas. Utilizando los resultados del análisis, la empresa establecía qué cliente insatisfecho precisaba que se le devolviera la llamada y, en relación con esto, analizaba automáticamente, entre otros aspectos, el estado emocional del interesado que llama, así como del empleado del servicio de atención al cliente, junto con otras características de la conversación. La finalidad de esta tecnología fue gestionar las quejas, controlar la calidad de las llamadas y del trabajo y aumentar la eficiencia de las personas trabajadoras. El asunto llegó a la Agencia Húngara de Protección de Datos por la queja de uno de los clientes. El cliente-denunciante formuló al banco algunas preguntas sobre la información publicada en la web de la empresa donde se hacía referencia al análisis de las grabaciones de sonido, sin embargo, no recibió una respuesta satisfactoria e instó el procedimiento ante la Autoridad nacional de Protección de Datos.

En la resolución de la Agencia Húngara de Protección de Datos se analiza el cumplimiento del principio de proporcionalidad. Según el banco, el software no incluía IA; no tomaba decisiones automatizadas y los resultados de su análisis podían ser utilizados exclusivamente con intervención e interpretación humana. Sin embargo, el análisis realizado por la Agencia de Protección de Datos concluye que el software utilizaba IA para tratar de forma automatizada datos personales, resultando, por un lado, en una lista de llamadas en el orden en que deben ser devuelto, y, por otro lado, las emociones recono-

cidas y las características de grabación de voz asociadas con cada llamada (por ejemplo, la duración de las pausas).

Es especialmente interesante la conclusión alcanzada por la Agencia de Protección de Datos cuando analiza la posibilidad de aplicar la protección reforzada de los datos biométricos recogida en el RGPD. A juicio de la Agencia, no se cumplen los requisitos en el caso concreto para ser datos biométricos. Según los hechos explorados del caso, el análisis de voz genera datos, pero estos datos no permiten identificar al titular de los datos, por lo que está ausente la condición de datos biométricos. Con base en esta información, los empleados de la entidad bancaria podían decidir a quién devolver la llamada del servicio de atención al cliente para abordar la insatisfacción. Si bien el software no está diseñado para manejar quejas individuales, las quejas reportadas por teléfono son atendidas de alguna manera por el personal de atención al cliente, independientemente del funcionamiento del software.

Tampoco se aplica el artículo 21 del RGPD sobre las decisiones automatizadas. Esto se debe a que quedan excluidos del artículo 21 RGPD las decisiones automatizadas negativas, es decir, cuando los interesados no son seleccionados para ser llamados o no se reporta ningún error administrativo, por lo que en estos casos se toma una decisión negativa sin intervención humana. Y sobre las decisiones positivas, es decir, el grupo seleccionado de clientes sí había intervención humana. En efecto, se requiere la intervención humana para tomar medidas adicionales en el caso de personas seleccionadas por el software para ser llamados o de empleados para ser revisados, por lo que para estas personas se logra un impacto significativo, pero la condición para una decisión basada en procesos totalmente automatizados no se cumple en el caso concreto.

En definitiva, la Agencia de Protección de Datos aplica el régimen general de garantías en materia de protección de datos. Y a partir de este indica que las actividades de análisis de voz realizadas por el banco utilizando IA, en particular la evaluación de las emociones de los interesados, plantean en sí mismos problemas de protección de datos. Cuando se utilizan herramientas para examinar las características psicolingüísticas y los tonos emocionales del habla no es suficiente la existencia formal del consentimiento de la persona. La tecnología de priorización basada sobre el tratamiento del habla supone una invasión de la privacidad y conlleva el riesgo de que el interesado no sea capaz de reconocer en el momento de dar el consentimiento y evaluar la incidencia en sus derechos. La Agencia indica que la tecnología aplicada permite a la entidad financiera obtener datos de los que el cliente ni siquiera es consciente, por lo que el uso de dichas herramientas reduce la posición del interesado de ser sujeto del procedimiento a ser objeto de este.

Además, se plantean incumplimientos de los principios de proporcionalidad y transparencia. No se proporcionó información a los afectados en relación con el análisis de voz por parte de IA o el propósito de dicho procesamiento y, por lo tanto, no hubo derecho a oponerse al tratamiento realizado. El banco no consideró adecuadamente los intereses en juego. El problema debería haber sido establecer la adecuación y proporcionalidad para el propósito dado del procesamiento; en cambio, la evaluación de la empresa se basó únicamente en sus propios intereses. En realidad, no consideró la pro-

porcionalidad y la posición del afectado, menospreciando los riesgos significativos para los derechos fundamentales.

La Agencia considera que el argumento de la empresa de que podía realizar las actividades empleando menos personal no es en sí mismo una justificación proporcionada y adecuada para la supresión de los derechos fundamentales de los interesados y por el uso de una forma de procesamiento de datos que considera indeseable y que implica un alto riesgo, incluso si se garantizan los derechos adecuados de los interesados. La innovación sólo beneficia a las personas si es apropiada, eficaz y va acompañado de garantías fuertes.

Se insiste en la resolución en la finalidad para la que fueron recabados los datos. Si bien la grabación de la voz es un elemento inevitable en la actividad de atención al cliente, incluso obligatorio en caso de reclamaciones, si el banco desea realizar más operaciones de procesamiento con la voz para analizarlas de forma automatizada utilizando nuevas y no del todo conocidas tecnologías, también debe cumplir con el artículo 6 (4) de RGPD, ya que pretende procesar datos personales para una finalidad distinta para la finalidad para la que se recogieron los datos.

Resulta clave el apartado de la resolución referido a los trabajadores y su posición de debilidad contractual. Indica la Agencia que no se proporciona un sistema adecuado de garantías para los empleados que se encuentran en relación de subordinación y, por tanto, son más vulnerables que los terceros. Los análisis de las emociones, cuya eficacia sigue sin probarse y profunda y severamente limita su derecho a la libre determinación, no puede sustentarse de manera razonable en el caso de los empleados. Dado que los empleados también están sujetos explícitamente a las normas relacionadas con el desempeño en el lugar de trabajo sobre la elaboración de perfiles según el artículo 4 (4) del RGPD, un análisis exhaustivo de las reglas y garantías aplicables a este también es necesario con carácter previo al procesamiento de datos con una nueva tecnología y la empresa no lo tuvo presente. Si un controlador utiliza métodos innovadores y tecnologías menos conocidos, las expectativas son más altas que para las tecnologías clásicas, por lo que mayores garantías y la planificación cuidadosa también deben aplicarse en la supervisión de los empleados. La elaboración de perfiles, en particular el análisis de las emociones de los empleados plantea una serie de problemas legales y éticos. Problemas que no han sido identificados y abordados por el banco en el curso del tratamiento de datos. Expresa la Agencia que son dudosas las posibilidades reales de las personas trabajadoras para oponerse a este tratamiento si tenemos en cuenta la subordinación.

Sobre la base de los argumentos expuestos, la Agencia de Protección de Datos concluye que las prácticas de procesamiento de datos del banco en relación con el análisis de las grabaciones de voz por parte del servicio de atención al cliente suponen una vulneración de los artículos del RGPD 5 (1)(a), 6 (1) y 6 (4). De conformidad con el artículo 12 (1) del RGPD, el banco debe proporcionar a los interesados la información mínima información necesaria para comprender el tratamiento de forma concisa y comprensible, de forma que los interesados sean al menos conscientes de la naturaleza básica del tratamiento. Esta información no se proporcionó por el banco y los clientes no podían sospechar que su voz se analizara automáticamente, y tampoco podía razonablemente

esperar que le devolvieran la llamada sin solicitarlo, entre otras cosas, por el tono de su voz. Por todo ello, se impuso al banco la multa de 670.000 € y le obligó a suspender el análisis de emociones con fundamento en los artículos 12, 24 y 25 RGPD.

Si se produjera un supuesto como el descrito, tras la aprobación del RIA, se aplicaría el primer nivel de riesgo y, por tanto, se trataría de un sistema de IA prohibido en el ámbito laboral al no concurrir las excepciones explicadas sobre los motivos médicos o de seguridad. Por lo tanto, este tipo de prácticas en las empresas están prohibidas.

El segundo de los impactos se refiere a una posible vertiente preventiva del RIA. La reciente aprobación del RIA ha supuesto un hito en el ámbito laboral pero también en el terreno de la seguridad y salud en el trabajo. Las referencias reiteradas a la salud y seguridad en el nuevo reglamento comunitario plantean algunos interrogantes que invitan a la reflexión: ¿Es el RIA una norma de prevención de riesgos laborales?; ¿qué impacto tiene el RIA sobre la Ley española de Prevención de Riesgos Laborales?

El sistema normativo de la prevención de riesgos laborales se compone de normas que tienen orígenes distintos. La cláusula de apertura del sistema que se recoge en el artículo 1º de la Ley de Prevención de Riesgos Laborales (Ley 31/1995, de 8 noviembre, de Prevención de Riesgos Laborales), pone de relieve la procedencia múltiple de las normas de prevención. El tenor literal del artículo 1º de la LPRL dice así: «La normativa sobre prevención de riesgos laborales está constituida por la presente Ley, sus disposiciones de desarrollo o complementarias y cuantas otras normas, legales o convencionales, contengan prescripciones relativas a la adopción de medidas preventivas en el ámbito laboral susceptibles de producirlas en dicho ámbito». Esto es, el sistema se nutre de normas específicamente preventivas, pero también de normas y reglas externas[28].

Como se ha dicho, el nuevo RIA realiza numerosas referencias a la salud y seguridad haciendo alusión expresa al ámbito laboral. Lo que nos permite plantearnos si en realidad esta norma se ha sumado al bloque normativo de la prevención de riesgos laborales. A diferencia de las Directivas comunitarias que han sido abundantes en materia de prevención de riesgos laborales y que precisan de una norma nacional de transposición, el RIA es una norma comunitaria de directa aplicación en España.

Como se ha anticipado, el RIA introduce prohibiciones respecto de determinados sistemas de IA y prevé obligaciones de cierta exigencia para las empresas y entidades públicas que empleen sistemas de IA de alto riesgo. Con la aprobación del RIA se incrementa la protección de los datos de salud de las personas trabajadoras. Si bien no se alteran de forma explícita las reglas y garantías de tratamiento de la salud de las personas empleadas previstas en el RGPD y en la LPRL, se producen avances en el RIA al incluir, con carácter general, la prohibición de los sistemas automatizados de reconocimiento de emociones de las personas trabajadoras (nivel 1 de riesgo) y su admisión condicionada en los niveles 2, 3 y 4 de riesgo.

En relación con las obligaciones de transparencia para las empresas que utilicen los sistemas automatizados de emociones en los niveles de riesgo permitidos, si adopta-

[28] Vid. mi libro A.B. Muñoz Ruiz, *Sistema normativo de la prevención de riesgos laborales*, Lex Nova, Valladolid, 2009.

mos un enfoque preventivo la empresa tendría además la obligación de incorporar los sistemas de IA de alto riesgo en la evaluación de riesgos laborales cuando su uso pueda producir al trabajador un efecto perjudicial considerable en su salud.

En definitiva, sería deseable un mayor desarrollo de la normativa de IA que clarifique cómo afecta la nueva normativa comunitaria a las empresas en el ámbito de la prevención de riesgos laborales. Este desarrollo podría tomar la forma de reforma de algunos preceptos de la Ley de Prevención de Riesgos Laborales y también podría ser útil la aprobación de una guía de actuación para empresas y sindicatos por parte del Instituto Nacional de Seguridad y Salud del Trabajo.

# Analítica de personas y discriminación algorítmica en procesos de selección y contratación

## People analytics and algorithmic discrimination in selection processes

Anna Ginès i Fabrellas
*Profesora Titular de Derecho del Trabajo*
*Universitat Ramon Llull, Esade*
ORCID ID: 0000-0001-6313-8355

doi: 10.20318/labos.2024.9034

*Resumen:* La utilización de sistemas de inteligencia artificial para adoptar decisiones automatizadas, que se originó en el ámbito de las plataformas digitales, se ha extendido a los procesos de selección y contratación de personas. Muchas empresas y portales de búsqueda de empleo incorporan técnicas de analítica de personas para adoptar decisiones de selección y contratación de forma automatizada. Si bien los sistemas de inteligencia artificial pueden simplificar enormemente estos procesos, generan riesgos sobre los derechos fundamentales de las personas, incluyendo situaciones de discriminación dada la existencia de sesgos en dichos sistemas. El presente artículo analiza el uso de sistemas de analítica de personas y decisión automatizada en el ámbito de la selección y contratación de personas, con el fin de identificar la existencia de sesgos en el uso de esta tecnología y analizar los retos que plantea su tratamiento jurídico.
*Palabras clave:* Analítica de personas, proceso de selección, sesgo, algoritmo, inteligencia artificial, discriminación algorítmica.

*Abstract:* The use of artificial intelligence systems to adopt automated decisions, which started in the field of platform work, has extended to selection and hiring processes. Many companies and job search portals incorporate people analytics techniques to adopt automated selection and hiring decisions. Although artificial intelligence systems can greatly simplify these processes, they can also generate risks for peoples' fundamental rights, including situations of discrimination due to the existence of biases in these systems. This article analyses the use of people analytics and automated decision systems in the field of selection and hiring of people, to identify the existence of biases in the use of this technology and determine the challenges it poses to its legal treatment.
*Keywords:* People analytics, selection process, bias, algorithm, artificial intelligence, algorithmic discrimination

[*]El presente artículo se ha realizado en el contexto del proyecto Proyecto "Digitalización del trabajo justa, equitativa y transparente" (DigitalWORK) financiado por FEDER UE - Ministerio de Ciencia, Innovación y Universidades - Agencia Estatal de Investigación - Proyecto PID2023-146944NB-I00.

## 1. Introducción

El uso de algoritmos y sistemas de inteligencia artificial para la adopción de decisiones automatizadas en materia laboral surgió en el ámbito de las plataformas digitales[1], si bien se ha extendido en los últimos años a empresas de la economía tradicional. Son muchas las empresas que utilizan tecnología inteligente para adoptar decisiones de contratación, asignación de tareas, determinación de salarios e, incluso, despidos[2].

En concreto, la utilización de sistemas de inteligencia artificial para adoptar decisiones de forma automatizada se ha extendido a los procesos de selección y contratación de personas. Muchas empresas y portales de búsqueda de empleo incorporan técnicas de analítica de personas (*people analytics*, según su denominación en inglés) para adoptar decisiones de selección y contratación de forma automatizada[3].

Las empresas emplean sistemas de inteligencia artificial que utilizan test de personalidad o inteligencia, rastreos de redes sociales[4], sistemas de reconocimiento facial o juegos para evaluar a las personas candidatas[5]. El sistema elabora un perfil profesional de la persona candidata y determina su mayor o menor probabilidad de encajar en el puesto de trabajo[6]. Un proceso de selección con miles de personas candidatas puede simplificarse enormemente mediante un sistema de inteligencia artificial que identifica a aquellas personas que mejor pueden encajar en el puesto de trabajo.

Sin embargo, los sistemas de inteligencia artificial plantean retos para los derechos fundamentales a la intimidad, protección de datos personales, seguridad y salud e igualdad y no discriminación. Tras analizar brevemente el impacto del uso de sistemas algorítmicos en el ámbito laboral sobre los derechos fundamentales de las personas, el artículo se centra en analizad el impacto sobre el derecho fundamental a la igualdad y no dis-

---

[1] Ginès i Fabrellas, Anna, *El trabajo en plataformas digitales. Nuevas formas de precariedad laboral*, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2021, p. 154.

[2] O'Neil, Cathy, Weapons of Math Destruction. How Big Data increases inequality and threatens democracy, Penguin Books, Reino Unido, 2016, p. 105 y ss.

[3] Ver Raub, McKenzie, "Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices", *Arkansas Law Review*, vol. 71, nº 2, 2018, p. 529-570; Kulkarni, Swatee y Che, Xiangdong, "Intelligent Software Tools for Recruiting", *Journal of International Technology and Information Management*, vol 28, nº 2, 2019, p. 6-7.

Según un estudio realizado a nivel mundial, el uso de técnicas de analítica de personas incrementó del 10% en 2016 al 39% en 2020, si bien existen importantes diferencias a nivel geográfico (Mercer, *Global Talent Trends 2020*, 2020, p. 38 (https://www.mercer.com/content/dam/mercer/attachments/private/global-talent-trends-2020-report.pdf)).

[4] Cheong, Marc, Lederman, Reeva, Mcloughney, Aidan, Njoto, Sheila, Ruppanner, Leah y Wirth, Anthony, "Gender Occupational Sorting: The role of Artificial Intelligence in Exacerbating Human Bias in STEM Employment", *CIS & Policy Lab*, The University of Melbourne, 26.6.2020, p. 11.

[5] Schellmann, Hilke, The Algorithm. How AI decides who gets hired, monitored, promoted & fired & why we need to fight back now, Hachette Books, Nova York, 2024.

[6] Cheong, Marc, Lederman, Reeva, Mcloughney, Aidan, Njoto, Sheila, Ruppanner, Leah y Wirth, Anthony, *Ethical implications of AI bias as a result of workforce gender imbalance*. Universidad de Melbourne, 2020, p. 11.

criminación que tienen los sistemas de decisión automatizada como consecuencia de la reproducción de sesgos y estereotipos de género, raza, orientación sexual, discapacidad, etc.[7]. La tecnología inteligente no solo no elimina las desigualdades y discriminaciones existentes en nuestra sociedad, sino que las reproduce, las sistematiza y las magnifica[8].

En este contexto, el presente artículo analiza el uso de sistemas de analítica de personas o decisión automatizada en el ámbito de la selección y contratación de personas, con el fin de identificar la existencia de sesgos en el uso de esta tecnología y determinar los retos que plantea para su tratamiento jurídico.

## 2. La emergencia de los algoritmos en la relación laboral

El uso de algoritmos y sistemas de inteligencia artificial para la adopción de decisiones automatizadas en el ámbito laboral apareció de la mano del trabajo en plataformas digitales[9]. Aunque el trabajo en plataformas ha centrado el debate en torno a la naturaleza jurídica de la relación de las personas que prestan servicios[10], es innegable que las plataformas digitales han actuado como verdaderas pioneras en el uso de algoritmos para adoptar decisiones automatizadas[11] de asignación de pedidos, franjas horarias o despidos –eufemísticamente llamado desconexión de la plataforma–.

En atención a variables como la localización, la puntuación obtenida por las personas clientas, el medio de transporte utilizado, el número de horas de conexión, el número de pedidos aceptados y rechazados o la mayor disponibilidad en horas de alta demanda, plataformas como Uber, Glovo, Deliveroo o UberEats han adoptado deci-

[7] Para otras publicaciones referentes a la discriminación algorítmica, ver Ginès i Fabrellas, Anna, "Algoritmos sesgados en el trabajo. Consideraciones entorno a su tratamiento jurídico", *Trabajo y Derecho*, nº 19, 2024, p. 1-37; y Ginès i Fabrellas, Anna, "Sesgos discriminatorios en la automatización de decisiones en el ámbito laboral: evidencias de la práctica", en Rivas Vallejo, Pilar (Directora), *Discriminación algorítmica en el ámbito laboral: perspectiva de género e intervención*, Thomson Reuters Aranzadi, 2022, p. 295-331.

[8] Deva, Surya, "Addressing the gender bias in artificial intelligence and Automation", *Open Global Rights*, 10.4.2020 (disponible en: https://www.openglobalrights.org/addressing-gender-bias-in-artificial-intelligence-and-automation/).

[9] Ginès i Fabrellas, Anna, El trabajo en plataformas digitales, op. cit., p. 154.

[10] Para un análisis detallado del trabajo en plataformas, ver Ginès i Fabrellas, Anna, *El trabajo en plataformas digitales*, *op. cit*.

Ver, asimismo, entre otras muchas publicaciones, Rogers, Brishen, "The Social Costs of Uber", The University of Chicago Law Review Dialogue, vol. 82, 2015, p. 85-102; de Stefano, Valerio, "The rise of the «just-in-time workforce»: On-demand work, crowdwork and labour protection in the «gig-economy»", Conditions of Work and Employment Series, nº 71, Organización Internacional del Trabajo, 2016; Prassl, Jeremias, *Humans as a service. The promise and perils of work in the gig economy*, Oxford University Press, Nueva York, 2018; Choudary, Sangeet Paul, "The architecture of digital labour platforms: policy recommendations on platform design for worker well-being", *ILO Future of Work Research Paper Series*, nº 3, 2018, p. 1-49.

[11] Adams-Prassl, Jeremias, "What if your boss was an algorithm? Economic Incentives, Legal Challenges, and the Rise of Artificial Intelligence at Work", *Comparative Labor Law and Policy Journal*, vol. 41, nº 1, 2019, p. 12 (versión electrónica).

siones de asignación de servicios o franjas horarias entre las personas trabajadoras[12]. Las personas con más horas de conexión en la plataforma, más servicios aceptados y realizados y mayor disponibilidad en horas de alta demanda son premiadas por el algoritmo, mientras que aquellas con menos horas de conexión, menor disponibilidad o mayor número de servicios rechazados son penalizados con menor acceso a servicios o preferencia para escoger franja horaria. Plataformas de tareas online, como Upwork, Fiverr o Amazon Mechanical Turk, también utilizan sistemas de gestión algorítmica del trabajo[13]. Estas plataformas analizan el número de servicios realizados, la rapidez en la ejecución de las tareas o el nivel de satisfacción de las personas clientas para determinar el acceso de las personas a tareas de mejor calidad. Mediante estrategias de gamificación, la actividad pasada en la plataforma es utilizada para determinar el acceso de las personas a nuevos niveles, como si de un vídeo juego se tratara.

Las plataformas digitales también han sido pioneras en el uso de sistemas de decisión automatizada para decisiones de despido. Por ejemplo, Uber desactivaba de la plataforma de forma automática a aquellas personas que registraban una puntuación inferior a 4,5[14]. Sin embargo, empresas de la economía tradicional han empezado a utilizar tecnología inteligente también para decisiones de extinción. Así, Amazon rastrea de forma milimétrica la actividad laboral de las personas trabajadoras, midiendo el número y velocidad de cajas empaquetadas, y genera avisos automáticos o, incluso, despide de forma automatizada aquellas personas que quedan por debajo de unos determinados umbrales de productividad[15]. La monitorización es tan intensiva, que las personas trabajadoras se ven incluso obligadas a renunciar a pausas para ir al baño o comer para mantener su trabajo[16].

El uso de fórmulas de dirección algorítmica del trabajo, donde la actividad pasada en la plataforma es utilizada para determinar la actividad futura, es una manifestación de la subordinación propia de la relación laboral[17], como así ha sido confirmado por el Tri-

[12] Rosenblat, Alex, *Uberland. How algorithms are rewriting the rules of work*, University of California Press, Oakland (Estados Unidos), 2018; Ginès i Fabrellas, Anna, *El trabajo en plataformas digitales, op. cit.*, p. 102.

Es importante, no obstante, tener en cuenta la naturaleza cambiante del funcionamiento de las plataformas digitales, adaptándose a las exigencias del mercado o la evolución de la jurisprudencia referente a los indicios de laboralidad en el trabajo en plataformas, si bien la esencia de la dirección algorítmica del trabajo se mantiene como característica definitoria del trabajo en plataformas.

[13] Bergvall-Kåreborn, Birgitta y Howcroft, Debra, "Amazon Mechanical Turk and the commodification of labor", New Technology, Work and Employment, vol. 29, nº 3, 2014, p. 213-223; Ginès i Fabrellas, Anna, *El trabajo en plataformas digitales*, *op. cit.*, p. 107-108.

[14] Ginès i Fabrellas, Anna, *El trabajo en plataformas digitales, op. cit.*, p. 113.

[15] Lecher, Colin, "How Amazon automatically tracks and fires warehouse workers for "productivity"", *The Verge*, 25.4.2019 (https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations).

Todos los vínculos a páginas web del presente artículo han sido verificados a fecha de 6.10.2024.

[16] Liao, Shannon, "Amazon warehouse workers skip bathroom breaks to keep their jobs, says report", *The Verge*, 16.4.2018 (https://www.theverge.com/2018/4/16/17243026/amazon-warehouse-jobs-worker-conditions-bathroom-breaks).

[17] Ginès i Fabrellas, Anna, El trabajo en plataformas digitales, op. cit., p. 103.

bunal Supremo[18] o como se incluye en la presunción de laboralidad de la disposición adicional vigesimotercera del Estatuto de los Trabajadores[19] introducida por la "Ley Rider"[20].

La dirección algorítmica del trabajo es tan intrínseca de la organización del trabajo en plataformas digitales que la Directiva del Parlamento Europeo y del Consejo relativa a la mejora de las condiciones laborales en el trabajo en plataformas digitales aprobada en abril de 2024 incluye un Capítulo III específicamente dedicado a la gestión algorítmica del trabajo. Esta regulación, según establece la exposición de motivos, tiene como objeto proteger las condiciones laborales de las personas que trabajan en plataformas y promover la transparencia, la equidad y la rendición de cuentas en el uso de sistemas algorítmicos[21].

Aunque empezó en el ámbito de las plataformas digitales, cada vez son más las empresas de la economía tradicional que utilizan algoritmos y sistemas de inteligencia artificial para adoptar decisión de gestión de personas automatizadas, tales como selección de personas y contratación, distribución de tareas, determinación de horarios, fijación de salarios, promociones o despidos[22].

La gestión algorítmica del trabajo es presentada como una oportunidad para que las empresas puedan mejor su productividad y competitividad. El uso de sistemas de inteligencia artificial permite adoptar decisiones de gestión de personas –ya sea contratación, promoción o despido– de forma mucho más rápida y efectiva[23]. Además, se alega que el uso de sistemas de decisión automatizada permite eliminar posibles errores o, incluso, sesgos inconscientes en materia de género, raza, apariencia física, etc. de las personas humanas a la hora de tomar decisiones. Es decir, la inteligencia artificial se presenta como una oportunidad para las empresas para tomar decisiones matemáticamente objetivas y basadas exclusivamente en méritos[24].

No obstante, la gestión algorítmica del trabajo plantea algunos riesgos y desafíos urgentes relacionados con el respeto a los derechos fundamentales de las personas trabajadoras. La mayor problemática que plantea la gestión algorítmica del trabajo es la po-

---

[18] STS, 4ª, 25.9.2020 (rec. núm. 4746/2019).

[19] Ginès i Fabrellas, Anna, "Disposición addicional 23. Presunción de laboralidad en el ámbito de las plataformas digitales de reparto", en del Rey Gunater, Salvador (Director), *Estatuto de los Trabajadores. Comentado y con jurisprudencia,* La Ley, 4ª edició, Madrid, 2022, p. 2027-2038.

[20] Real Decreto-ley 9/2021, de 11 de mayo, por el que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales y, posteriormente convalidado por la Ley 12/2021, de 28 de septiembre.

[21] Ver Avogaro, Matteo, "La dirección algorítmica en la propuesta de Directiva sobre el trabajo en plataformas: un avance parcial entre la dimensión individual y colectiva", en Ginès i Fabrellas, Anna (Directora), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 231-265.

[22] O'Neil, Cathy, Weapons of Math Destruction, op. cit., p. 105 y ss.

[23] Kuncel, Nathan R., Ones, Deniz S. y Klieger, David M., "In Hiring, Algorithms Beat Instinct", *Harvard Business Review*, Mayo 2014 (https://hbr.org/2014/05/in-hiring-algorithms-beat-instinct); Kim, Pauline, "Big Data and Artificial Intelligence: New Challenges for Workplace Equality", *University of Louisville Law Review*, vol. 57, 2019, p. 316; Kulkarni, Swatee y Che, Xiangdong, "Intelligent Software Tools for Recruiting", *op. cit.*, p. 13.

[24] Kulkarni, Swatee y Che, Xiangdong, "Intelligent Software Tools for Recruiting", *op. cit.*, p. 8.

tencial afectación a los derechos fundamentales de las personas. El Parlamento Europeo, en su resolución de 14 de marzo de 2017 referente a las implicaciones de los macrodatos en los Derechos fundamentales[25], indicó que el uso de macrodatos para el tratamiento automatizado mediante algoritmos o el uso de sistemas de inteligencia artificial genera "*riesgos significativos, concretamente en lo que se refiere a la protección de derechos fundamentales como el derecho a la privacidad, la protección y la seguridad de los datos, además de la libertad de expresión y la no discriminación, garantizados por la Carta de los Derechos Fundamentales y la legislación de la Unión*". No es de extrañar que el uso de sistemas de inteligencia artificial para la adopción de decisiones de selección o contratación de personas o para la determinación de condiciones laborales, promociones, la extinción del contrato, asignación de tareas y monitorización o evaluación del comportamiento o desempeño de las personas trabajadoras se haya calificado como de alto riesgo en la Unión Europea (ver artículo 6 en relación con el Anexo III del Reglamento de Inteligencia Artificial[26]).

Los derechos fundamentales a la intimidad y protección de datos personales resultan amenazados por la proliferación de sistemas de inteligencia artificial. La utilización de modelos de decisión automatizada requiere de grandes volúmenes de datos, de distinto origen y tipología[27], para su entrenamiento[28]. Los sistemas de inteligencia artificial son entrenados en base a grandes volúmenes de datos con el fin de identificar conexiones o patrones estadísticos en los datos y, en atención a estos, hacer predicciones de comportamiento o preferencias de las personas[29], que posteriormente son utilizadas como referencia para tomar decisiones en el futuro.

El nivel de extracción de datos que requiere el entrenamiento de sistemas de inteligencia artificial supone un incremento de la vigilancia y control social, conduciéndonos a lo que Shoshana ZUBOFF califica como capitalismo de vigilancia[30]. El incremento de

[25] Resolución del Parlamento Europeo de 14 de marzo de 2017 sobre las implicaciones de los macrodatos en los Derechos fundamentales: privacidad, protección de datos, no discriminación, Seguridad y aplicación de la ley (2016/2225(INI)).

[26] Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).

[27] Los datos pueden provenir, por ejemplo, de redes sociales, historial de búsquedas en internet, aplicaciones informáticas, sistemas de reconocimiento facial, sistemas de geolocalización, etc.

En el ámbito laboral, los datos utilizados pueden provenir de currículums, entrevistas de trabajo, la actividad digital de las personas trabajadoras, sensores digitales o *wearables* utilizados en el ámbito laboral (por ejemplo, sistemas de geolocalización), aplicaciones informáticas (por ejemplo, de salud o *fitness*), etc. Ver, en relación con esta cuestión, Adams-Prassl, Jeremias, "When Your Boss Comes Homes", *C4E The Future of Work in the Age of Automation and AI*, 2020, p. 5 (https://c4ejournal.net/2020/07/05/jeremias-adams-prassl-when-your-boss-comes-home-2020-c4ej-xxxx-symposium/).

[28] Raub, McKenzie, "Bots, Bias and Big Data…", *op. cit.*, p. 532; Cheong *et al.*, "Gender Occupational Sorting…", *op. cit.*, p. 6.

[29] Kim, Pauline, "Big Data and Artificial Intelligence…", *op. cit.*, p. 317.

[30] Zuboff, Shoshana, The age of surveillance capitalism. The fight for a human future at the new frontier of power, Profile Books, Reino Unido, 2019.

la vigilancia y control se produce también en la empresa, donde la utilización de sistemas de decisión automatizada requiere acceder a mucha información de las personas trabajadoras, en muchas ocasiones afectando incluso a su esfera privada[31]. Se produce un incremento sin precedentes de las capacidades de vigilancia y control de la empresa[32], dada a la elevada información de las personas trabajadoras disponible en las redes sociales o en internet. Sin embargo, lo que genera una verdadera ventana indiscreta para la empresa[33], es la capacidad de los modelos predictivos de acceder a todavía más información personal de las personas trabajadoras. La información más inofensiva puede utilizarse para predecir datos personales altamente sensibles. Así, por ejemplo, la actividad en redes sociales puede ser utilizada para predecir datos personales como el sexo, orientación sexual, origen racial, opiniones políticas, edad, nivel de inteligencia, uso de sustancias adictivas o situación parental[34]. Más allá, los modelos predictivos pueden incluso hacer predicciones respecto de personas que no comparten información en redes sociales, basándose en la premisa que las personas que comparten determinadas características tienen unas mismas preferencias o comportamiento[35].

El nivel de extracción y procesamiento de datos requerido para el entrenamiento de sistemas de decisión automatizada puede afectar también al derecho fundamental a la protección de datos personales. A pesar de la estricta regulación que ofrece el Reglamento General de Protección de Datos[36] (RGPD, en adelante), el elevado volumen de datos que requiere el entrenamiento de sistemas de inteligencia artificial parece incompatible con el respeto de principios esenciales, como el principio de minimización de datos *ex* artículo 5.1.c) RGPD, que exige procesar solamente los datos estrictamente necesarios para su finalidad, o el principio de limitación de la finalidad *ex* artículo 5.1.b) RGPD, que impide el tratamiento ulterior de datos por finalidad incompatible con la original. Incluso cuando los datos son recolectados y almacenados de forma anonimizada, pueden existir riesgos de vulneración del derecho a la protección de datos personales. Aunque los principios de protección de datos no aplican a la información anónima[37], en ocasiones,

---

En sentido similar, Carissa Véliz califica el fenómeno como economía de vigilancia (Véliz, Carissa, *Privacy is power. Why and how you should take back control of your data*, Transworld publishers, Londres, 2020, p. 3).

[31] Adams-Prassl, Jeremias, "When Your Boss Comes Homes", *op. cit.*, p. 5.

[32] de Stefano, Valerio, "Algorithmic Bosses and How to Tame Them", *C4E The Future of Work in the Age of Automation and AI*, 2020, p. 14 (https://c4ejournal.net/2020/07/05/valerio-de-stefano-algorithmic-bosses-and-how-to-tame-them-2020-c4ej-xxx/).

[33] Ifeoma Ajunwa, Kate Crawford y Jason Schultz lo califican como vigilancia infinita (Ajunwa, Ifeoma, Crawford, Kate y Schultz, Jason, "Limitless Worker Surveillance", *California Law Review*, vol. 105, nº 3, 2017, p. 735-776).

[34] Kosinski, Michal, Stillwell, David y Graepel, Thore, "Private traits and attributes are predictable from digital records of human behavior", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 110, nº 15, 2013, p. 5803.

[35] Véliz, Carissa, *Privacy is power*, *op. cit.*, p. 88-96.

[36] Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo 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 y por el que se deroga la Directiva 95/46/CE.

[37] Según el considerando 26 RGPD, "los principios de protección de datos no deben aplicarse a la infor-

combinados con otras bases de datos, pueden resultar identificadas o identificables las personas físicas[38]. Además, es posible que la empresa no cumpla con sus obligaciones de información ante la adopción de decisiones automatizadas o elaboración de perfiles *ex* artículos 13, 14 y 15 RGPD en relación con el artículo 22 RGPD[39]. Y, en consecuencia, que las personas no sepan que son objeto de decisiones automatizadas ni qué información la empresa está infiriendo[40].

El derecho a la seguridad y salud de las personas trabajadoras también resulta amenazado por la introducción de sistemas de inteligencia artificial[41]. Desde mi punto de vista, la inteligencia artificial puede resultar una gran aliada en la mejora de la seguridad y salud en el trabajo. Así, por ejemplo, mediante la utilización de sensores digitales es posible reducir o, incluso, eliminar situaciones de riesgo; o, por poner otro ejemplo, el análisis de datos referentes a siniestralidad laboral, contingencias profesionales o absentismo laboral puede revelar información interesante en materia de riesgos laborales que, bien utilizada, puede mejorar la política de prevención de la empresa[42].

Sin embargo, la utilización de sistemas de inteligencia artificial para la adopción de decisiones automatizadas puede tener efectos corrosivos sobre la seguridad y salud de las personas trabajadoras. Cuando los sistemas de inteligencia artificial son utilizados para la medición de la productividad pueden traducirse en una intensificación del trabajo, que puede dar lugar a un incremento de accidentes, riesgos físicos y/o riesgos psicosociales. En este sentido, existe evidencia del impacto de los sistemas de decisión automatizada sobre la salud de las personas en el ámbito del trabajo en plataformas[43], así como también en empresas de la economía tradicional[44].

---

mación 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".

[38] Adams-Prassl, Jeremias, "When Your Boss Comes Homes", *op. cit.*, p. 5.

[39] Ginès i Fabrellas, Anna, "Decisiones automatizadas y elaboración de perfiles en el ámbito laboral y su potencial impacto discriminatorio", en Ginès i Fabrellas, Anna (Directora), *Algoritmos, Inteligencia Artificial y Relación Laboral*, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2023, p. 173-229.

[40] Wachter, Sandra y Mittelstadt, Brent, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI", *Columbia Business Law Review*, vol. 2, 2019, p. 500.

[41] Dzieza, Josh, "How hard will the robots make us work?", *The Verge*, 27.2.2020 (https://www.theverge.com/2020/2/27/21155254/automation-robots-unemployment-jobs-vs-human-google-amazon).

[42] Luque Parra, Manel, "IA y seguridad y salud laboral: la dicotomía entre ser un gran aliado productivo y un "riesgo laboral emergente"", en Ginès i Fabrellas, Anna (Directora), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 305-334.

[43] Garben, Sacha, *Protecting workers in the online platform economy: an overview of regulatory and policy developments in the EU*, European Risk Observatory Discussion paper, European Agency for Safety and Health at Work, Luxemburgo, 2017, p. 24-28; Huws, Ursula, Spencer, Neil H., Syrdal, Dag S. y Holts, Kaire, *Work in the European Gig Economy. Research results from the UK, Sweden, Germany, Austria, The Netherlands, Switzerland and Italy*, Foundation for European Progressive Studies, UNI Europa y University of Hertfordshire, 2017, p. 47.

[44] Amazon, por ejemplo, tiene unos elevados de siniestralidad laboral claramente superiores a los registradas en otras empresas del sector, especialmente durante las campañas de Black Friday o Cyber Monday (Hamilton, Isobel Asher y Cain, Áine, "Amazon warehouse employees speak out about the "brutal" reality of working during the holidays, when 60-hour weeks are mandatory and ambulance calls are common", *In-*

Téngase en cuenta que los efectos negativos sobre la seguridad y salud de las personas trabajadoras aparecen, incluso, cuando el sistema de decisión automatizada no se utiliza de forma expresa como mecanismo de control. Los sistemas de inteligencia artificial para adoptar decisiones automatizadas –por ejemplo, de asignación de tareas, salarios, promociones, etc.– requieren de un elevado control y evaluación continua de la actividad de las personas trabajadoras. Existe una monitorización intrínseca en los sistemas de decisión automatizada, que puede provocar una intensificación del trabajo que, como se ha apuntado, puede incrementar los riesgos físicos y psicosociales.

Además, el uso de sistemas de inteligencia artificial en el ámbito laboral genera riesgos psicosociales intrínsecos asociados al miedo a perder el trabajo, a sufrir discriminación algorítmica, a la invasión de espacios de intimidad por la monitorización constante, a la inseguridad que genera no conocer cómo funciona el algoritmo y, por tanto, como comportarse para obtener buenas valoraciones o a la pérdida de autonomía o capacidad para realizar el trabajo[45].

La utilización de sistemas de inteligencia artificial y de decisión automatizada también supone un riesgo para el respeto del derecho fundamental a la igualdad y no discriminación, al que se destina el presente artículo. Los sistemas de inteligencia artificial incluyen sesgos y estereotipos de género, raza, orientación sexual, discapacidad, etc., que pueden generar situaciones de verdadera discriminación[46]. La literatura científica ha evidenciado que la tecnología inteligente no solo no elimina por arte de magia las desigualdades y discriminaciones existentes en nuestras sociedades, sino que las reproduce, las sistematiza y las magnifica[47].

En este contexto, el objeto del presente trabajo es analizar el uso de sistemas de analítica de personas y decisión automatizada en el ámbito de la selección y contratación de personas, con el fin de identificar la existencia de sesgos en el uso de esta tecnología y analizar los retos que plantea para su tratamiento jurídico.

---

*sider*, 19.2.2019 (https://www.businessinsider.com/amazon-employees-describe-peak-2019-2); Peters, Jay, "Internal documents show automated Amazon warehouses have higher injury rates", *The Verge*, 29.9.2020 (https://www.theverge.com/2020/9/29/21493752/amazon-warehouses-robots-higher-injury-rates-report-reveal). En este contexto, el estado de California aprobó una ley aplicable a centros de distribución de almacenes, que obliga a las empresas a informar a las personas trabajadoras y la autoridad laboral sobre los criterios de productividad exigidos e impide criterios de productividad incompatibles con los adecuados periodos de descanso, comida o pausas para ir al baño -y, expresamente menciona, el tiempo razonable para ir y volver del baño (Ley AB-701 de 22 septiembre de 2021, disponible en: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB701).

[45] Moore, Phoebe V., "Making Algorithmic Management safe for workers: new regulation is needed", 28.7.2023 (disponible en: https://phoebevmoore.wordpress.com/2023/07/29/making-algorithmic-management-safe-for-workers-new-regulation-is-needed/?s=09).

[46] Ver Ginès i Fabrellas, Anna, "Algoritmos sesgados en el trabajo…", *op. cit*.

[47] Deva, Surya, "Addressing the gender bias...", *op. cit*.

## 3. La analítica de personas en procesos de selección y contratación

La utilización de sistemas de inteligencia artificial para adoptar decisiones de forma automatizada se ha extendido en los últimos años a los procesos de selección y contratación de personas. Muchas empresas y portales de búsqueda de empleo incorporan técnicas de analítica de personas (*people analytics*, según su denominación en inglés) para adoptar decisiones de selección y contratación de forma automatizada[48].

En el ámbito de la selección de personas, las empresas emplean sistemas de inteligencia artificial que utilizan entrevistas virtuales mediante sistemas de reconocimiento facial o tono de voz, evaluación automatizada de CVs, juegos o rastreos de redes sociales[49] para evaluar a las personas candidatas a un puesto de trabajo[50]. El sistema elabora un perfil profesional de la persona y determina su mayor o menor probabilidad de encajar en el puesto de trabajo[51].

Cuando una empresa quiere contratar a alguien para ocupar un puesto de trabajo, le puede interesar conocer distintos aspectos de la persona como, por ejemplo, si es buena trabajadora, si tiene capacidad de liderazgo, si trabaja bien en equipo, si acepta las críticas, si tiene iniciativa, etc. Sin embargo, esta información, si bien relevante para tomar la decisión, generalmente es desconocida por parte de la empresa. Por este motivo, algunas empresas recurren a sistemas de inteligencia artificial que analizan la información disponible para predecir las características personales que a la empresa realmente le interesa saber[52]. Es decir, se utiliza información de la persona candidata disponible en su CV, redes sociales, entrevista de trabajo, etc., para configurar un perfil profesional de la persona, predecir como actuará y se comportará y adoptar la decisión de selección o contratación en base a esta predicción.

En primer lugar, algunas empresas utilizan sistemas de reconocimiento facial o tono de voz para analizar a las personas candidatas durante una entrevista virtual de trabajo y, en función de la información identificada por el sistema, configurar un perfil profesional de la persona, y así ranquearla y compararla con las demás personas candidatas al puesto de trabajo para tomar la decisión de selección[53].

Uno de los ejemplos más conocidos es el de la empresa HireVue, que desarrolló un software de evaluación de personas en un proceso de selección[54]. El software utiliza un sistema de reconocimiento facial que analiza a las personas durante una entrevista virtual,

---

[48] Kulkarni, Swatee y Che, Xiangdong, "Intelligent Software Tools for Recruiting", *op. cit.*, p. 6-7; Raub, McKenzie, "Bots, Bias and Big Data…", *op. cit.*

[49] Cheong *et al.*, "Gender Occupational Sorting…", *op. cit.*, p. 11.

[50] Kulkarni, Swatee y Che, Xiangdong, "Intelligent Software Tools for Recruiting", *op. cit.*, p. 5-8; Schellmann, Hilke, *The Algorithm*, *op. cit.*

[51] Cheong et al., Ethical implications of AI bias as a result of workforce gender imbalance, op. cit., p. 11.

[52] O'neil, Cathy, Weapons of Math Destruction, op. cit., p. 17.

[53] Schellmann, Hilke, *The Algorithm*, *op. cit.*, p. 83 y ss.

[54] Harwell, Drew, "A face-scanning algorithm increasingly decides whether you deserve the job", *The Washington Post*, 6.11.2019 (https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/).

evaluando las palabras utilizadas para elaborar un perfil profesional de la persona, midiendo sus capacidades y competencias, con la finalidad de identificar a la persona que mejor encaja en el puesto de trabajo ofertado[55]. Anteriormente, el sistema también analizaba los movimientos faciales y tono de voz de las personas entrevistadas, si bien la empresa dejó de analizar estos elementos –aunque la empresa no lo ha reconocido, no existía evidencia científica detrás de esta práctica[56], además de las consideraciones éticas de determinar el acceso al empleo en base a características genéticas de las personas, como su tono de voz–[57].

Empresas de búsqueda de empleo también utilizan *chatbots* para chatear con las personas candidatas en un proceso de selección[58]. Por ejemplo, StepStone adquirió el sistema de inteligencia artificial Mya para integrarlo en su plataforma y así "*descubrir más sobre las preferencias, habilidades e intereses de quienes buscan empleo y aumentar el número y la calidad de las solicitudes de empleo coincidentes*"[59]. El sistema saca conclusiones de las conversaciones que mantiene con las personas que buscan empleo y les manda sugerencias de puestos de trabajo que podrían encajar con sus intereses y perfil.

En segundo lugar, algunas empresas utilizan sistemas de inteligencia artificial para evaluar de forma automatizada los CV de las personas candidatas a un puesto de trabajo[60]. La empresa proporciona al algoritmo los CV de las personas que considera "*top performers*" en la empresa; es decir, aquellas personas que considera han demostrado mejor rendimiento en la empresa. El algoritmo, mediante un sistema de aprendizaje automático, identifica aquellas características que comparten estas personas y, posteriormente, las busca en los CV de las personas candidatas al puesto de trabajo. La lógica de estos sistemas es que aquellas personas que comparten unas determinadas características -por ejemplo, formación, *alma mater*, experiencia profesional previa, etc.- es probable que se comportaran de la misma manera. Por ejemplo, si el algoritmo identifica que de entre las personas "*top performers*" hay un porcentaje significativo que han estudiado en una determinada universidad, buscará a personas candidatas que hayan estudiado en esa misma universidad, prediciendo que estas tendrán un buen rendimiento profesional.

En tercer lugar, otros sistemas de inteligencia artificial muy utilizados en la actualidad en procesos de selección son los juegos. Algunas empresas han desarrollado sistemas de inteligencia artificial que evalúan a las personas en atención a su desempeño en determinados videojuegos[61]. Empresas como Pymetrics o Knack han desarrollado juegos de ordenador que, utilizados en procesos de selección, permiten medir variables como

---

[55] Raub, McKenzie, "Bots, Bias and Big Data…", *op. cit.*, p. 538.

[56] La investigación de Hilke Schellmann demuestra como el sistema llega a puntuar muy bien cuando la respuesta consiste en leer al azar una entrada de la Wikipedia en idiomas distintos a los empleados por el software (Schellmann, Hilke, *The Algorithm*, *op. cit.*, p. 123-124).

[57] Dattner, Ben, Chamorro-Premuzic, Tomas, Buchband, Richard y Schettler, Lucinda, "The Legal and Ethical Implications of Using AI in Hiring", *Harvard Business Review*, 25.4.2019.

[58] Raub, McKenzie, "Bots, Bias and Big Data…", *op. cit.*, p. 537.

[59] Ver https://www.stepstone.de/ueber-stepstone/press/stepstone-expands-autonomous-matching-acquires-us-conversational-ai-technology-mya/

[60] Schellmann, Hilke, *The Algorithm*, *op. cit.*, p. 1 y ss.

[61] Schellmann, Hilke, *The Algorithm*, *op. cit.*, p. 51 y ss.

la atención, asertividad, capacidad de decisión, esfuerzo, emoción, foco, generosidad, aprendizaje o tolerancia al riesgo de las personas[62]. El modelo mide el rendimiento en estos juegos de las mejores personas empleadas por la empresa y lo utiliza como estándar de referencia para evaluar a las personas candidatas[63].

Un juego de Pymetrics es uno donde van apareciendo globos en la pantalla y a medida que el globo se va hinchando, se va acumulando dinero[64]. Cuando el globo explota, se pierde el dinero acumulado. El objetivo del juego es maximizar el dinero recaudado y, por tanto, recolectar el dinero antes de que explote el globo. Al rato de jugar, las personas empiezan a observar que los globos naranjas y amarillos explotan antes, por tanto, hay que recolectar el dinero rápido, mientras que los azules tardan más en explotar y, por tanto, hay que esperar para poder maximizar el dinero recaudado. Este juego, por ejemplo, pretende predecir la capacidad y rapidez de aprendizaje, aversión al riesgo, generosidad o atención de las personas.

En cuarto lugar, otro sistema utilizado en procesos de selección es el escaneo de redes sociales[65]. Mediante sistemas de inteligencia artificial es posible analizar la interacción de las personas en redes sociales y, en base a esta información, elaborar un perfil profesional de la persona. La empresa Deep-Sense ofrecía servicios a otras empresas de escanear las redes sociales de las personas candidatas a un puesto de trabajo para predecir su encaje en la empresa, personalidad y comportamiento.

Un proceso de selección con miles de personas candidatas puede simplificarse enormemente mediante un sistema de inteligencia artificial que identifica a aquellas personas que mejor pueden encajar en el puesto de trabajo. La empresa puede realizar entrevistas individuales con las personas mejor evaluadas por el modelo, simplificándose enormemente el procedimiento, recudiendo asimismo los costes empresariales[66].

Sin embargo, los sistemas de inteligencia artificial pueden estar complicado y dificultando la toma de decisiones en un proceso de selección. Aunque pretenden mejorar la eficiencia y la velocidad en la selección de personas, pueden provocar el efecto contrario como consecuencia del incremento de la información a procesar. Un sistema de inteligencia artificial puede mejorar un proceso de selección con miles de personas candidatas, al permitir un escaneo automatizado de currículums o entrevistas. Si bien, simultáneamente lo está complicando, dado el incremento de solicitudes que estos sistemas permiten[67].

Más allá, como se analiza en el siguiente apartado, los sistemas de inteligencia artificial incluyen sesgos y estereotipos de género, raza, orientación sexual, discapacidad, etc., que se reproducen en los sistemas de decisión automatizada generando verdaderas situaciones de discriminación.

---

[62] Andrews, Lori y Bucher, Hannah, "Automating Discrimination: AI hiring practices and gender inequality", *Cardozo Law Review*, vol 44, nº 1, 2022, p. 185-186.

[63] Schellmann, Hilke, *The Algorithm*, *op. cit.*, p. 63.

[64] Schellmann, Hilke, "Auditors are testing hiring algorithms for bias, but there's no easy fix", *MIT Technology Review*, 11.2.2021.

[65] Schellmann, Hilke, *The Algorithm*, *op. cit.*, p. 29 y ss.

[66] Kulkarni, Swatee y Che, Xiangdong, "Intelligent Software Tools for Recruiting", *op. cit.*, p. 13.

[67] Schellmann, Hilke, *The Algorithm*, *op. cit.*, p. xiii.

## 4. Discriminación algorítmica en procesos de selección y contratación de personas

El uso de sistemas de inteligencia artificial en procesos de selección puede generar situaciones de discriminación algorítmica[68], entendida como aquella situación de discriminación generada por el uso de algoritmos o sistemas de inteligencia artificial, por producir un trato desfavorable no justificado por razón de sexo, raza, religión, edad, identidad sexual u otras casusas de discriminación prohibidas por la Constitución o la ley.

La discriminación algorítmica, como se analiza a continuación, puede encontrar su origen en la existencia de sesgos en las variables utilizadas por el algoritmo para tomar decisiones, en la base de datos sobre la que se ha entrenado el algoritmo o en las variables proxy o correlaciones identificadas por el algoritmo[69].

En primer lugar, la discriminación algorítmica tiene su origen en la existencia de sesgos en las variables que utiliza el algoritmo para tomar decisiones cuando el algoritmo directa o indirectamente utiliza algunas de las variables protegidas de sexo, raza, edad, identidad sexual, etc. para tomar decisiones.

A modo de ejemplo, la utilización de técnicas de publicidad segmentada para ofertas de trabajo podría generar una situación de discriminación prohibida cuando se utilice –ya sea de forma directa o indirecta– una variable de discriminación prohibida para seleccionar o excluir a determinados colectivos de la oferta de trabajo. El sistema de inteligencia artificial puede identificar a aquellas personas o colectivos en la plataforma o en redes sociales a quien dirigir la oferta de trabajo[70] (por ejemplo, a personas con una determinada formación). Y esta práctica, *a priori*, es legal y legítima, por cuanto la empresa pretender mejor dirigir las ofertas de trabajo a los colectivos potencialmente interesados. Incluso la práctica comúnmente utilizada en portales de empleo de dirigir ofertas de empleo a personas que comparten determinadas características con otras (*lookalike audience*, según denominación en inglés)[71], resultaría *a priori* lícita.

No obstante, estas prácticas pueden generar una discriminación cuando la segmentación se realice –directa o indirectamente– por alguna de las causas de discriminación prohibidas y, por tanto, se incluya o excluya a colectivos en atención a causas de

[68] Dattner, Ben *et al.*, "The Legal and Ethical Implications of Using AI in Hiring", *op. cit.*

[69] La clasificación utilizada en el presente artículo respecto el origen de la discriminación algorítmica resulta, desde mi punto de vista, especialmente adecuada a efectos de analizar su tratamiento jurídico. Sin embargo, para distintas clasificaciones referentes al origen o la causa de la discriminación algorítmica, ver, por ejemplo, Costa, Allan, Cheung, Chris y Langenkamp, Max, "Hiring Fairly in the Age of Algorithms", *Research Paper Human-Computer Interaction*, Cornell University, 2020, p. 11-18; Unceta, Irene, "Notas para un aprendizaje automático justo", en Ginès i Fabrellas, Anna (Directora), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 95-99; UNESCO, *Challenging systematic Prejudices: an investigation into Gender Bias in Large Language Models*, 2024.

[70] Kim, Pauline, "Big Data and Artificial Intelligence…", *op. cit.*, p. 316.

[71] Moreno Cáliz, Susana, "Análisis del comportamiento de las plataformas de captación, selección y contratación de trabajadores que emplean algoritmos para la adopción de decisiones: evidencias", en Rivas Vallejo, Pilar (Directora), *Discriminación algorítmica en el ámbito laboral: perspectiva de género e intervención*, Thomson Reuters Aranzadi, 2022, p. 221.

discriminación prohibidas. Las redes sociales –por ejemplo, Facebook– tienen el potencial para discriminar la visualización de ofertas de trabajo por origen género, racial, edad u otras causas de discriminación prohibidas[72]. Este es el caso de la empresa T-Mobile, que restringió las visualizaciones de una oferta de trabajo en Facebook a personas entre 18 y 30 años[73]. Por poner otro ejemplo, LinkedIn identificó que el algoritmo mostraba ofertas de empleo a aquellas personas que había identificado tenían más probabilidad de contactar con la empresa o postularse para el puesto, lo que generaba una discriminación por razón de género, por cuanto generalmente los hombres tienen una actitud más agresiva en la búsqueda de empleo[74] y tienden a presentarse a ofertas de trabajo incluso si no cumplen todos los requerimientos[75].

Otro ejemplo de discriminación derivado de sesgos en las variables utilizadas por el algoritmo lo encontramos en el uso de juegos en los procesos de selección. Por un lado, debe cuestionarse la adecuación de estas técnicas de evaluación basadas en juegos y su aplicación en procesos de selección, por cuanto parece poco ético que un juego de ordenador determine la mayor o menor probabilidad de acceder a un empleo[76]; personas menos habituadas o familiarizadas con juegos de ordenador pueden verse penalizadas, a pesar de tener un buen perfil profesional.

Pero, más allá, el uso de juegos en procesos de selección puede tener un efecto discriminatorio al haberse evidenciado diferencias entre hombres y mujeres en la ejecución de dichos juegos, así como por razón de edad[77]. No existen diferencias significativas entre el número de mujeres y hombres usuarias de videojuegos; a pesar de la menor presencia de personajes de videojuegos femeninos[78], se estima que las mujeres representan la mitad de las personas que juegan a videojuegos a nivel mundial[79]. Sin embargo, como se ha apuntado, sí se han detectado diferencias de género y edad en el comportamiento en estos juegos utilizados en procesos de selección, lo que puede generar una situación discriminatoria. También puede tener un impacto discriminatorio por razón de discapacidad o enfermedad, por cuanto muchos de estos juegos miden la rapidez de reacción. Al medir el tiempo de compleción de una determinada tarea, es posible que el sistema penalice a personas con una determinada enfermedad o discapacidad que carecen de la

[72] Kim, Pauline, "Big Data and Artificial Intelligence…", *op. cit.*, p. 317.

[73] Ver decisión de *United States District Court, Northern District of California, San Jose Division* en el asunto Bradley *et al.* v. T-Mobile US, Inc. e*t al.* (caso nº 17-cv-07232-BLF), si bien el caso fue desestimado por cuestiones procesales.

VWall, Sheridan y Schellmann, Hilke, "LinkedIn's job-matching AI was biased. The company's solution? More AI", *MIT Technology Review*, 23.6.2021.

VNicks, Leonie, Gesiarz, Filip, Valencia, Lourdes, Hardy, Tim y Lohmann, Johannes, *Gender differences in response to requirements in job adverts*, The Behavioural insights team, Government Equalities Office, 2022.

[76] Andrews, Lori y Bucher, Hannah, "Automating Discrimination…", *op. cit.*, p. 189.

[77] Melchers, Klaus G. y Basch, Johannes M., "Fair play? Sex-, age-, and job-related correlates of performance in a computer-based simulation game", *International Journal of Selection and Assessment*, nº 30, p. 48-61.

[78] Criado Perez, Caroline, Invisible Women. Exposing data bias in a world designed for men, Vintage, Londres, 2019, p. 12.

[79] Chen, Vickie, "Leveling Up the Gaming Gender Field", *Forbes*, 24.8.2023 (https://www.forbes.com/sites/forbesbusinesscouncil/2023/08/24/leveling-up-the-gaming-gender-gap/).

rapidez exigida por el juego, a pesar de que la rapidez no tiene ninguna relación con el puesto de trabajo ofertado[80].

Los sistemas de inteligencia artificial que analizan los movimientos faciales o tono de voz de las personas en una entrevista de trabajo también pueden tener un efecto discriminatorio al confundir errores de lenguaje de personas no nativas[81], migradas o refugiadas, así como diferencias culturales de expresión con un mal perfil profesional[82]. Más allá, es interesante recordar la investigación de Joy Buolamwini y Timnit Gebru[83], que evidenció la existencia de sesgos raciales en los principales sistemas de reconocimiento facial, registrando porcentajes de error significativamente más elevados para las mujeres negras que para los hombres blancos (34,7% vs. 0,8% en el software de IBM).

Los riesgos de los sistemas de reconocimiento facial sin duda han motivado su inclusión en el listado de sistemas prohibidos en el Reglamento de Inteligencia Artificial. El artículo 5.1 prohíbe, entre otros, los sistemas de inteligencia artificial dirigidos a inferir emociones en el ámbito de la relación laboral, salvo cuando se utilice con fines médicos o de seguridad (artículo 5.1.f), así como tampoco los que permiten inferir la raza, opiniones políticas, afiliación sindical, creencias religiosas o filosóficas, vida sexual u orientación sexual (artículo 5.1g))[84]. Téngase en cuenta, no obstante, que no se prohíbe el uso de todo sistema de reconocimiento facial en el ámbito laboral, sino solamente aquellos dirigidos a inferir emociones o a predecir información sensible de las personas.

En segundo lugar, la discriminación algorítmica puede obedecer a la existencia de sesgos en la base de datos utilizada para entrenar el algoritmo[85]. Los algoritmos requieren de grandes volúmenes de datos para su entrenamiento; se les proporciona grandes volúmenes de datos para que puedan identificar conexiones y patrones estadísticos dentro de la base de datos y generar un modelo matemático que permita hacer predicciones y, en base a estas, tomar decisiones de forma automatizada[86]. Por impresionante que parezca –y, sin duda, lo es–, el problema es que el modelo matemático configurado sobre una base de datos sesgada incorporará dicho sesgo y se traducirá en las predicciones y decisiones adoptadas[87].

---

[80] Mercader Uguina, Jesús R., *Algoritmos e inteligencia artificial en el derecho del trabajo*, Tirant lo Blanch, Valencia, 2022, p. 76-77.

[81] O'neil, Cathy, *Weapons of Math Destruction*, *op. cit.*, p. 116.

[82] Feldman, Lisa, Adolphs, Ralph, Marsella, Stacy, Martinez, Aleix M. y Pollak, Seth D., "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements", *Psychological Science in the Public Interest*, vol. 20, nº 1, 2019, p. 1-68.

[83] Buolamwini, Joy y Gebru, Timnit, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", *Conference of Fairness, Accountability, and Transparency. Proceedings of Machine Learning Research*, vol. 81, 2018, p. 6.

[84] Véase Muñoz Ruiz, Ana Belén, Biometría y sistemas automatizados de reconocimiento de emociones: implicaciones jurídico-laborales, Tirant lo Blanch, Valencia, 2023.

[85] Cheong, Marc *et al.*, *Ethical implications of AI bias*, *op. cit.*, p. 11.

[86] Raub, McKenzie, "Bots, Bias and Big Data…", *op. cit.*, p. 533.

[87] Zou, James, "Removing gender bias from algorithms", *The Conversation*, 26.9.2019 (https://theconversation.com/removing-gender-bias-from-algorithms-64721).

La existencia de sesgos en los datos de entrenamiento es la principal causa de discriminación algorítmica[88] y, seguramente, la más difícil de abordar. Los sistemas de decisión automatizada son entrenados con los datos disponibles; por ejemplo, datos reales referentes a decisiones o situaciones pasadas, que generalmente incluyen sesgos y discriminaciones pasadas[89]. Cuando los distintos grupos no están adecuadamente representados en la base de datos de entrenamiento, se magnifican las características del grupo dominante, que se toman como referencia para tomar decisiones, lo que se denomina sesgo de representación[90].

Un ejemplo de discriminación algorítmica por sesgos en la base de datos de entrenamiento en el contexto de un proceso de selección es el conocido caso del sistema de inteligencia artificial creado por Amazon para la selección de personas[91]. El modelo tenía como finalidad identificar el perfil profesional que mejor encaja en la empresa en atención a las contrataciones de la empresa de los últimos 10 años. La intención de la empresa era utilizar este sistema en procesos de contratación para predecir aquellas personas que mejor encajarían en la empresa. No obstante, dado que durante el periodo de referencia las contrataciones habían sido mayoritariamente de hombres, el algoritmo identificó que los hombres encajan mejor en la empresa y, en consecuencia, descartaba automáticamente los CVs que contenían la palabra "mujer" o que identificaba provenían de mujeres. Ni tan siquiera se evaluó el impacto discriminatorio por raza del sistema, seguramente también alarmante[92].

En tercer lugar, la discriminación algorítmica puede encontrar su origen en sesgos en las variables proxy o correlaciones estadísticas identificadas por el algoritmo[93], denominado sesgo por correlación o discriminación por poxy. Como se ha apuntado anteriormente, los algoritmos para la elaboración de perfiles utilizan la información disponible sobre la persona (variables proxy), para hacer predicciones sobre sus características, comportamiento o aptitudes[94]. Por ejemplo, en el marco de un proceso de selección, el modelo utiliza la información disponible que se correlaciona con un buen perfil profesional (por

---

[88] Costa, Allan *et al.*, "Hiring Fairly in the Age of Algorithms", *op. cit.*, p. 11.

[89] Mcfarland, Daniel y Mcfarland, H. Richard, "Big Data and the danger of being precisely inaccurate", *Big Data & Society*, 2015, p. 1.

[90] Unceta, Irene, "Notas para un aprendizaje automático justo", *op. cit.*, p. 96.

Los sesgos en la base de datos también pueden provenir de errores o imprecisiones en la recolección de datos (sesgos de medida) o sesgos derivados de la combinación de datos de grupos heterogéneos que conlleva que, si bien los distintos grupos están igual representados, el modelo no es capaz de representar adecuadamente a ninguno (sesgos de agregación). Ver Crawford, Kate, "The Hidden Biases in Big Data", *Harvard Business Review*, 1.4.2013 (https://hbr.org/2013/04/the-hidden-biases-in-big-data); Unceta, Irene, "Notas para un aprendizaje automático justo", *op. cit.*, p. 98.

[91] Vincent, James, "Amazon reportedly scraps internal AI recruiting tool that was biased against women", *The Verge*, 10.10.2018 (disponible en: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G); Dastin, Jeffrey, "Amazon scraps secret AI recruiting tool that showed bias against women", *Reuters*, 11.10.2018 (disponible en: https://www.reuters.com/article/idUSKCN1MK0AG/).

[92] Benjamin, Ruha, *Race after technology*, Polity Press, Medford (EUA), 2019, p. 143.

[93] Cheong, Marc et al., Ethical implications of AI bias, op. cit., p. 11.

[94] O'neil, Cathy, Weapons of Math Destruction, op. cit., p. 17.

ejemplo, formación, experiencia profesional previa, entrevista de trabajo, etc.), para hacer una predicción sobre la persona (capacidad de trabajo, liderazgo, trabajo en equipo, etc.) y, en base a esto, tomar decisiones de selección y contratación. Cuando la variable proxy o la correlación estadística identificada por el modelo está indirectamente relacionada con una variable de discriminación prohibida, puede generar un efecto discriminatorio prohibido.

A modo de ejemplo, hay empresas que utilizan como variable proxy para predecir la probabilidad de que la persona permanezca más tiempo en la empresa, la distancia entre el trabajo y el domicilio[95], por cuanto las personas que residen más lejos del trabajo y, por tanto, dedican más tiempo de desplazamiento, tienen más probabilidad de aceptar un trabajo más cercano a su residencia. Aunque existe un interés legítimo de la empresa de querer contratar y formar a aquellas personas que vayan a permanecer más tiempo en la empresa, esta variable aparentemente inofensiva puede tener un efecto discriminatorio. Los astronómicos precios de la vivienda en el centro de muchas ciudades, es una variable que puede generar una discriminación por capacidad económica u origen racial[96], por cuanto las personas que residen en la periferia estadísticamente son aquellas con menos recursos económicos y/o personas migradas.

También es interesante el ejemplo de la *start-up* Gild, que, en un artículo publicado en *The Atlantic* en 2013, reveló una curiosa correlación que había identificado su algoritmo[97]. Gild era una *start-up* que ofrecía servicios de búsqueda de talento a empresas tecnológicas, elaborando un perfil profesional de las personas programadoras. Mediante distintas variables, el modelo otorgaba una calificación sobre 100 a las personas, ranqueándolas en su base de datos de más de 6 millones de personas programadoras. La novedad del modelo es que cuantificaba el "capital social" de la persona, entendido como su integración en la comunidad digital, asociando mayor integración con una mejor puntuación. El modelo identificó una correlación estadística entre las habilidades de codificar de una persona y la participación en foros de programación, como Stack Overflow, o la visita a una concreta página web de manga japonés. Además de extremadamente curioso, esta variable aparentemente inofensiva puede generar un efecto discriminatorio si tenemos en cuenta el contenido altamente sexual que caracteriza el manga japonés, que puede no resultar atractivo para las mujeres[98].

La empresa admitía en dicho artículo que la conexión entre visitar una página de manga japonés y buenas habilidades para codificar no es una relación de causalidad; no por visitar dicha página de manga japonés se desarrollan buenas habilidades de programación. Se trata simplemente de una correlación estadística curiosa e interesante identificada por el algoritmo. Sin embargo, dado que los algoritmos no distinguen entre correlación

[95] O'neil, Cathy, Weapons of Math Destruction, op. cit., p. 119.

[96] Kim, Pauline, "Big Data and Artificial Intelligence…", *op. cit.*, p. 317.

[97] Peck, Don, "They're Watching You at Work", *The Atlantic*, diciembre 2013 (https://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/).

[98] O'neil, Cathy, *Weapons of Math Destruction*, *op. cit.*, p. 120-121; Criado Perez, Caroline, *Invisible Women*, *op. cit.*, p. 107.

y causación[99], pueden tomar decisiones en base a correlaciones estadísticas, aunque nada tengan que ver con la decisión adoptada. Si bien en el caso de Gild esta variable era una entre muchas otras que el algoritmo utilizaba para elaborar el perfil profesional de la persona, sí que actuaba como "*nudge*" o empujón para subir la cualificación de la persona.

Más allá, desde mi punto de vista, también puede tener un efecto discriminatorio utilizar como variable para predecir las habilidades de codificar de una persona su integración en la comunidad digital. El modelo de Gild analizaba las aportaciones realizadas en estos foros, calificando mejor aquellas respuestas con mayor popularidad y aceptación en el foro. Sin embargo, teniendo en cuenta la desigual distribución de tareas de cuidado entre mujeres y hombres, es posible que las mujeres programadoras no puedan pasar horas y horas en foros de programación[100]. Además, es importante tener en cuenta la elevada presencia de hombres en estos foros y la hostilidad que en ocasiones pasadas ha mostrado la comunidad digital hacia la incorporación de mujeres[101], que puede actuar como barrera de entrada para las mujeres programadoras.

En este ejemplo se observa el sesgo de correlación, al utilizar como variables proxy para predecir características personales variables que generan un efecto desfavorable hacia un colectivo protegido; en este caso, el colectivo de mujeres. Es más, aunque nada se mencione en el artículo publicado en *The Atlantic*, es posible que este sesgo se encuentre reforzado también por un sesgo en la base de datos de entrenamiento del algoritmo, presumiblemente integrada mayoritariamente por hombres, dada la menor presencia de mujeres en este sector. Las mujeres solamente representan el 20% de las personas que desarrollan roles técnicos en la industria de la inteligencia artificial, el 12% de las investigadoras en inteligencia artificial y el 6% de desarrolladoras de software profesionales[102].

## 5. La opacidad algorítmica como principal reto para el tratamiento jurídico de la discriminación algorítmica

La discriminación algorítmica podrá encauzarse, desde mi punto de vista, en la actual doctrina antidiscriminatoria[103], sin que sea necesaria la creación de categorías jurídicas

[99] Costa, Allan *et al.*, "Hiring Fairly in the Age of Algorithms", *op. cit.*, p. 11.

[100] O'neil, Cathy, *Weapons of Math Destruction*, *op. cit.*, p. 120-121; Criado Perez, Caroline, *Invisible Women*, *op. cit.*, p. 107.

[101] Lorenz, Taylor y Browning, Kellen, "Dozens of Women in Gaming Speak Out About Sexism and Harassment", *The New York Time*, 23.6.2020 (https://www.nytimes.com/2020/06/23/style/women-gaming-streaming-harassment-sexism-twitch.html); Jankowicz, Nina, "Online Harassment Towards Women Is Getting Even More Insidious", *Wired*, 28.1.2021 (https://www.wired.com/story/online-harassment-toward-women-getting-more-insidious/).

[102] UNESCO, "Women's access to and participation in technological developments" (https://www.unesco.org/en/artificial-intelligence/gender-equality).

[103] Ginès i Fabrellas, Anna, "Sesgos discriminatorios…", *op. cit.*, p. 312.

Ver también en este sentido Preciado Domenech, Carlos Hugo, "Algoritmos y discriminación en la relación laboral", *Jurisdicción Social*, nº 223, 2021, p. 17; Rivas Vallejo, Pilar, *La aplicación de la Inteli-*

nuevas[104]. A mi entender, la discriminación algorítmica generalmente caerá dentro de la discriminación indirecta, por tratarse de una disposición, criterio o práctica aparentemente neutra que pone a personas de un colectivo protegido en una desventaja particular con respecto a las personas del otro grupo, "*salvo que dicha disposición, criterio o práctica puedan justificarse objetivamente en atención a una finalidad legítima y que los medios para alcanzar dicha finalidad sean necesarios y adecuados*" (artículo 6.2 de la Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres (LOI, en adelante))[105]. Cuando el algoritmo adopte decisiones en base a variables que generen una desventaja particular sobre un colectivo protegido, estaremos ante un caso de discriminación indirecta, salvo que la empresa pueda probar de forma objetiva y proporcional una finalidad legítima para atender a dicha variable en el proceso de toma de decisión.

Así, por ejemplo, constituiría un caso de discriminación indirecta un algoritmo utilizado en un proceso de selección que penalice, mediante un peor perfil profesional, a aquellas personas con interrupciones en sus carreras profesionales, por cuanto genera una desventaja particular a las personas que han tenido interrupciones laborales por motivos de una enfermedad, discapacidad o para el cuidado de hijos o familiares dependientes. También constituiría un caso de discriminación indirecta, a mi entender, el sistema de decisión automatizada que, en un proceso de selección, evalúa a las personas en base a su actuación en juegos de ordenador y analiza, por ejemplo, la rapidez de reacción, penalizando a aquellas personas con una determinada enfermedad o discapacidad.

Sin perjuicio de lo anterior, no es posible descartar la existencia de supuestos de discriminación directa por decisiones algorítmicas, cuando el algoritmo utilice alguna variable de discriminación prohibida para tomar decisiones. A modo de ejemplo, constituiría discriminación directa el uso de técnicas de publicidad segmentada en redes sociales que limitan la visualización de una oferta de trabajo a determinados grupos por razón de edad. Asimismo, desde mi punto de vista también constituiría discriminación directa el ejemplo apuntado del sistema de selección de personas de Amazon que automáticamente descartaba a las mujeres[106].

---

*gencia Artificial al trabajo y su impacto discriminatorio*, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2020, p. 303; Fernández García, Antonio, "Trabajo, algoritmos y discriminación", en Rodríguez-Piñero Royo, Miguel y Todolí Signes, Adrián (Directores), *Vigilancia y control en el Derecho del Trabajo Digital*, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2020, p. 524.

[104] En sentido contrario, no obstante, se posicionan algunas personas autoras que sí abogan a favor de la ampliación de la tutela antidiscriminatoria. Ver, en este sentido, Sáez Lara, Carmen, "El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación", *Temas Laborales*, nº 155, 2020, p. 49; Pérez del Prado, Daniel, *Derecho, economía y digitalización. El impacto de la inteligencia artificial, los algoritmos y la robótica sobre el empleo y las condiciones de trabajo*, Tirant lo Blanch, Valencia, 2023, p. 188; Todolí Signes, Adrián, *Algoritmos productivos y extractivos. Cómo regular la digitalización para mejorar el empleo e incentivar la innovación*, Aranzadi, Cizur Menor (Navarra), 2023, p. 73.

[105] Para un estudio más detallado referente al tratamiento jurídico de la discriminación algorítmica ver Ginès i Fabrellas, Anna, "Algoritmos sesgados en el trabajo. Consideraciones entorno a su tratamiento jurídico", *Trabajo y Derecho*, nº 19, 2024, p. 1-37.

[106] Adams-Prassl, Jeremias, Binns, Reuben y Kelly-Lyth, Aislinn, "Directly Discriminatory Algorithms", *The Modern Law Review*, vol. 86, nº 1, p. 167.

Si bien la discriminación algorítmica podrá tratarse mediante la actual tutela discriminatoria, su tratamiento jurídico genera retos derivados de la opacidad algorítmica[107]. La opacidad y falta de transparencia es una característica intrínseca de los algoritmos y sistemas de inteligencia artificial[108]. Por un lado, existe un interés empresarial a no ofrecer información sobre el funcionamiento del algoritmo porque, además de estar protegido por secreto empresarial[109], existe el interés de evitar que las personas puedan utilizar esta información para ganar al modelo[110]. Pero, más allá, los algoritmos más complejos son creados con técnicas de *machine learning* y *deep learning* que generan multitud de capas de decisión que, conocidos como algoritmos de "cajas negras", impiden que las decisiones puedan ser completamente explicadas[111].

Esta opacidad algorítmica entra en colisión con la necesidad de aportar indicios de discriminación. Como es bien sabido, el proceso de tutela de derechos fundamentales prevé que, "*una vez justificada la concurrencia de indicios de que se ha producido violación del derecho fundamental o libertad pública, corresponderá al demandado la aportación de una justificación objetiva y razonable, suficientemente probada, de las medidas adoptadas y de su proporcionalidad*" (artículo 181.2 la Ley 36/2011, de 10 de octubre, reguladora de la jurisdicción social (LRJS, en adelante)). En sede de discriminación indirecta, la evidencia estadística de desventaja particular hacia un colectivo protegido puede actuar como indicio para activar la inversión de la carga de la prueba[112].

No obstante, la información que podría actuar como indicio de discriminación no está generalmente disponible para las personas trabajadoras o la representación legal de la plantilla. Así, por ejemplo, información sobre las variables utilizadas por el modelo para tomar decisiones, las correlaciones estadísticas identificadas por el algoritmo o el efecto de las decisiones no es generalmente información disponible.

Es cierto que la actual normativa regula derechos de información a las personas sujetas a decisiones automatizadas. El artículo 22 RGPD parte de la existencia de una

---

[107] Castillo, Carlos, "Discriminación algorítmica. Aproximación conceptual", en Rivas Vallejo, Pilar (Directora), *Discriminación algorítmica en el ámbito laboral: perspectiva de género e intervención*, Thomson Reuters Aranzadi, 2022, p. 75.

[108] O'neil, Cathy, *Weapons of Math Destruction*, *op. cit.*, p. 28; Eubanks, Virginia, *Automating Inequality. How high-tech tools profile, police, and punish the poor*, Picador, Nueva York, 2019, p. 185; Costa, Allan *et al.*, "Hiring Fairly in the Age of Algorithms", *op. cit.*, p. 8.

[109] Kullmann, Miriam, "Platform Work, Algorithmic Decision-Making and EU Gender Equality Law", *International Journal of Comparative Labour Law and Industrial Relations*, vol. 34, nº 1, 2018, p. 15 (versión digital).

[110] O'neil, Cathy, Weapons of Math Destruction, op. cit., p. 28.

[111] de Laat, Paul B., "Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?", *Philosophy & Technology*, nº 31, 2017, p. 11 (versión digital); Cheong, Marc *et al.*, *Ethical implications of AI bias*, *op. cit.*, p. 14.

Es interesante apuntar también las voces que abogan por la utilización de sistemas de decisión automatizada más simples en beneficio de la explicabilidad (ver Rudin, Cynthia, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisiones and Use Interpretable Models Instead", *Nature Machine Intelligence*, vol. 1, 2019, p. 1-20 (versión digital)).

[112] STJUE 28.2.2013 (caso c-427/11, asunto *Kenny et al.*).

prohibición de decisiones automatizadas, incluyendo la elaboración de perfiles, con efectos jurídicos o similarmente significante[113]. No obstante, cuando dichas decisiones automatizadas sean admitidas por concurrir alguna de las excepciones previstas por el propio precepto (por ejemplo, la necesidad para la celebración o ejecución de un contrato, que podría ser aplicada en el ámbito laboral), se reconoce el derecho de las personas sujetas a dichas decisiones automatizadas a obtener "*información significativa sobre la lógica involucrada, así como la importancia y las consecuencias previstas de dicho procesamiento*" (artículos 13.2.f), 14.2.g) y 15.1.h) RGPD).

No obstante, como se argumenta a continuación, los derechos de información actualmente regulados son limitados y no permiten solucionar el problema que la opacidad algorítmica genera sobre el tratamiento jurídico de la discriminación derivadas de sistemas de inteligencia artificial.

En primer lugar, existe una indeterminación en cuanto al contenido de los derechos de información, por cuanto los artículos 13.2.f), 14.2.g) y 15.1.h) RGPD solamente se refieren a "*información significativa sobre la lógica involucrada*" sin mayor concreción. El Grupo de Trabajo del Artículo 29 ha interpretado que el derecho de información incluye el derecho a obtener información clara y simple sobre el funcionamiento del proceso de elaboración de perfiles o decisión automatizada, con el fin de entender los motivos de la decisión[114]. En consecuencia, parece seguro afirmar que debe proporcionarse información sobre (i) el uso de sistemas automatizados para tomar decisiones y las decisiones en las que se utiliza, (ii) las variables utilizadas y su posición relativa en la ecuación y (iii) las consecuencias que pueden derivarse para las personas trabajadoras.

Sin embargo, desde mi punto de vista, también se debe proporcionar información referente a la base de datos de entrenamiento del sistema, por cuanto es información esencial para entender la "*lógica*" del algoritmo. Así lo ha entendido también la Agencia Española de Protección de Datos[115] o el Ministerio de Trabajo y Economía Social[116]. Más allá, acceder a esta información puede resultar necesario para el tratamiento jurídico de una eventual discriminación algorítmica. Desde mi punto de vista, la existencia de sesgos prohibidos en la base de datos utilizada para entrenar el algoritmo debería admitirse como evidencia suficiente para apreciar indicios de discriminación y permitir la inversión de la carga de la prueba[117]. La existencia de sesgos en la base de datos de entrenamiento es

[113] Para un análisis detallado de la regulación en materia de decisiones automatizadas y los derechos de información ver Respecto de las obligaciones de información ante decisiones automatizadas en el ámbito laboral, ver Ginès i Fabrellas, A., "Decisiones automatizadas y elaboración de perfiles…", *op. cit.*

[114] Article 29 Data Protection Working Party, "Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679", 3.10.2017, p. 25.

[115] Agencia Española de Protección de Datos, "Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción", 2020.

[116] Ministerio de Trabajo y Economía Social, "Información algorítmica en el ámbito laboral. Guía práctica y herramienta sobre la obligación empresarial de información sobre el uso de algoritmos en el ámbito laboral", mayo 2022 (https://www.mites.gob.es/ficheros/ministerio/inicio_destacados/Guia_Algoritmos_ES.pdf).

[117] Ginès i Fabrellas, Anna, "Algoritmos sesgados en el trabajo…", *op. cit.*

la causa más habitual de discriminación algorítmica[118]; la literatura científica ha evidenciado que los algoritmos entrenados sobre bases de datos sesgadas reproducen en mayor medida incluso el sesgo presente en los datos de entrenamiento[119]. En cualquier caso, la evidencia estadística de sesgos en la base de datos de entrenamiento serviría simplemente como indicio, descartándose la existencia de discriminación si la empresa aportar "*una justificación objetiva y razonable, suficientemente probada, de las medidas adoptadas y de su proporcionalidad*" (artículo 181.2 LRJS).

La falta de referencia expresa sobre la obligación de proporcionar información sobre la base de datos de entrenamiento utilizada puede generar dudas en torno a su obligación y exigencia y, en consecuencia, dificultar el tratamiento jurídico de la discriminación algorítmica derivados de sesgos en la base de datos de entrenamiento. Concretar el contenido de la información que debe proporcionar la empresa a las personas trabajadoras sujetas a decisiones automatizadas *ex* artículo 22 RGPD es esencial a efectos de, entre otras, garantizar el acceso a la información necesaria para actuar como indicio de discriminación algorítmica.

En segundo lugar, el artículo 22 RGPD y los derechos de información *ex* artículos 13.2.f), 14.2.g) y 15.1.h) RGPD se refieren a decisiones íntegramente automatizadas sin intervención humana[120]. El artículo 22.1 RGPD es claro al establecer que las personas tienen "*derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él* [ella] *o le afecte significativamente de modo similar*". En consecuencia, no se reconocen derechos de información respecto de las decisiones semiautomatizadas en las que también hay intervención humana significativa[121]. En estos supuestos, las personas trabajadoras no tendrían acceso a información alguna. Téngase en cuenta que, cuando en un proceso de selección el sistema clasifica de forma automatizada a las personas en atención a su perfil profesional, se trata de una decisión íntegramente automatizada, aunque posteriormente la empresa proceda a realizar entrevistas personales con las mejor posicionadas; en la fase inicial del proceso de selección ha habido una decisión automatizada sin intervención humana y, en consecuencia, corresponde aplicar la regulación del artículo 22 RGPD.

---

[118] Costa, Allan *et al.*, "Hiring Fairly in the Age of Algorithms", *op. cit.*, p. 11.

[119] Cheong, Marc *et al.*, *Ethical implications of AI bias as a result of workforce gender imbalance*, *op. cit.*, p. 9; Bolukbasi, Tolga, Chang, Kai-Wei, Zou, James, Saligrama, Venkatesh y Kalai, Adam, "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings", *NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, p. 1-9; James Zou, "Removing gender bias from algorithms", *op. cit.*

[120] En relación con esta regulación, ver también Mercader Uguina, Jesús R., *Algoritmos e inteligencia artificial en el derecho digital del trabajo*, *op. cit.*, p. 159-164.

[121] Es importante destacar que la intervención humana debe ser significativa para descartar la aplicación del artículo 22 RGPD y los derechos de información asociados. Ver Article 29 Data Protection Working Party, "Guidelines on automated individual decision-making…", *op. cit.*, p. 21; Malgieri, Gianclaudio y Comandé, Giovanni, "Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation", *International Data Privacy Law*, vol. 7, nº 4, 2017, p. 14 (versión electrónica).

Esta limitación de la normativa europea de protección de datos parece haberse mitigado con la introducción en el ordenamiento jurídico-laboral español de una obligación específica de información a la representación legal de la plantilla de obtener información sobre los "*parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento de empleo, incluida la elaboración de perfiles*" (artículo 64.4.d) ET)[122]. Este derecho de información viene a suplir algunas de las limitaciones de los derechos de información incluidos en la normativa de protección de datos, por cuanto resulta de aplicación a toda forma de decisión automatizada, incluyendo también las que cuentan con intervención humana, y se reconoce a la representación legal de la plantilla[123]. Sin embargo, esta obligación de información tampoco garantiza el acceso a la información más relevante y necesaria para el tratamiento jurídico de la discriminación algorítmica, que, como se argumenta a continuación, es la información referente al impacto o efectos de las decisiones adoptadas por el algoritmo[124].

Efectivamente, en tercer lugar, no existe la obligación de información respecto del efecto de las decisiones tomadas por el algoritmo, que es la información más relevante para actuar como indicio de discriminación algorítmica; especialmente respecto de los algoritmos de caja negra o variables proxy aparentemente inofensivas. Como se ha apuntado anteriormente, existen algoritmos creados mediante complejas técnicas que, al tener multitud de capas de decisión, resultan poco explicables[125]. En consecuencia, es posible que respecto de estos algoritmos creados mediante técnicas de aprendizaje automático no sea posible acceder a la información sobre las variables utilizadas. Dificultando, en consecuencia, que las personas trabajadoras o la representación legal de la plantilla puedan apreciar que una determinada variable está generando un efecto discriminatorio. Téngase en cuenta que la discriminación algorítmica puede resultar menos evidente o aparente. Los sistemas de decisión automatizada, envueltos de un aura de objetividad y neutralidad, reducen la sensación de estar siendo discriminadas[126].

---

[122] En relación con los derechos colectivos de información, consulta y negociación de la representación legal de la plantilla en materia del uso de algoritmos y sistemas de inteligencia artificial, ver Pastor Martínez, Alberto, "Los derechos colectivos de información, consulta y negociación del uso de algoritmos y sistemas de Inteligencia Artificial", en Ginès i Fabrellas, Anna (Directora), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 335-361.

[123] Ministerio de Trabajo y Economía Social, "Información algorítmica en el ámbito laboral", *op. cit.*

[124] Ginès i Fabrellas, Anna, "El derecho a conocer el algoritmo: una oportunidad perdida de la "Ley Rider", *IUSLabor*, nº 3, 2021, p. 1-5.

[125] de Laat, Paul B., "Algorithmic Decision-Making Based on Machine Learning from Big Data…", *op. cit.*, p. 11; Cheong, Marc *et al.*, *Ethical implications of AI bias*, *op. cit.*, p. 14.

Es interesante apuntar también las voces que abogan por la utilización de sistemas de decisión automatizada más simples en beneficio de la explicabilidad (ver Rudin, Cynthia, "Stop Explaining Black Box Machine Learning Models…", *op. cit.*).

[126] Wachter, Sandra, Mittelsdtadt, Brent y Russell, Chris, "Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI", *Computer Law & Security Review*, vol. 41, 2021, p. 6 (version digital); Castillo, Carlos, "Discriminación algorítmica", *op. cit.*, p. 75.

En consecuencia, resulta esencial para apreciar si un sistema algorítmico genera una desventaja particular sobre un colectivo protegido conocer como el algoritmo está tomando decisiones respecto de este colectivo en comparación con otros. Por ejemplo, en un proceso de selección de personas, resultaría esencial tener una comparativa de los datos desagregados por sexo del número de solicitudes recibidas en relación con el número de personas admitidas en las distintas fases del proceso de selección para apreciar indicios de discriminación. Sin embargo, como se ha apuntado, esta información no se reconoce en la normativa actual.

El Reglamento de Inteligencia Artificial introduce nuevas obligaciones de información y transparencia de los sistemas de inteligencia artificial de alto riesgo, entre los que, como se ha apuntado anteriormente, se incluyen los utilizados en el ámbito laboral. El artículo 13 exige que los sistemas de inteligencia artificial de alto riesgo sean diseñados y desarrollados para garantizar transparencia e interpretabilidad en su funcionamiento, requiriéndose la elaboración de unas instrucciones de uso que incluyan información referente a, entre otras, las características, capacidades y limitaciones del sistema. A su vez, el artículo 26 del Reglamento de Inteligencia Artificial, dirigido a las entidades responsables del despliegue del sistema, entre las que se incluye la empresa que utiliza estos sistemas, establece la obligación de informar a las personas trabajadoras y a la representación legal de la plantilla de la utilización de sistemas de inteligencia artificial en el ámbito laboral. Más allá, el artículo 86 establece la obligación de la empresa de proporcionar una explicación clara y significativa del rol que ha jugado el sistema de inteligencia artificial en un proceso de decisión y los principales elementos utilizados para adoptar la decisión, en los supuestos de sistemas de alto riesgo con efectos legales o significativos y que puedan tener un impacto negativo sobre la salud, seguridad o derechos fundamentales de las personas.

Sin perjuicio de los nuevos derechos de información introducidos por el Reglamento de Inteligencia Artificial, desde mi punto de vista es importante avanzar hacia una regulación específica para el ámbito laboral del uso de sistemas de inteligencia artificial. Se trata todavía de una regulación parcial que no concreta la información que debe facilitarse y limita el derecho de información de la representación legal de la plantilla a información sobre el uso de sistemas de inteligencia artificial en el ámbito laboral, sin más información sobre las características del sistema.

En consecuencia, desde mi punto de vista es necesario reconocer obligaciones específicas y concretas de información del uso de algoritmos y sistemas de inteligencia artificial para la toma de todo tipo de decisiones automatizadas, incluidas las que cuentan con intervención humana; información dirigida a las personas trabajadoras y a la representación legal de la plantilla[127]. En concreto, en línea con la Guía de información algorítmica en el ámbito laboral publicada por el Ministerio de Trabajo y Economía Social[128], la empresa debe informar (i) acerca del uso de sistemas de inteligencia artificial para tomar decisiones en el ámbito laboral, identificando la tecnología utilizada, las concretas

[127] Para un análisis más detallado ver Ginès i Fabrellas, A., "Decisiones automatizadas y elaboración de perfiles…", *op. cit*.

[128] Ministerio de Trabajo y Economía Social, "Información algorítmica en el ámbito laboral", *op. cit*.

decisiones sujetas a sistemas de decisión automatizada y, en su caso, el grado de intervención humana cualificada en la decisión; (ii) el funcionamiento del sistema, incluyendo información sobre la tipología de perfiles que elabora el algoritmo y la asignación de la persona a un concreto perfil, las variables, parámetros y reglas utilizadas por el algoritmo en la toma de decisiones, los datos de entrenamiento y, en su caso, validación, utilizados y sus características, las métricas de precisión o error del modelo y las auditorías o evaluación de impacto realizada; y, finalmente, (iii) sobre las consecuencias que pueden derivarse de la decisión adoptada en términos de acceso al empleo, mantenimiento del empleo o determinación de condiciones laborales.

Además, desde mi punto de vista, es necesario incluir la obligación de realizar auditorías previas e independientes de los sistemas de decisión automatizados utilizados en el ámbito de la relación laboral[129] y la obligación de facilitar información sobre los resultados a la representación legal de la plantilla. En esta línea, es interesante apuntar que la Directiva de trabajo en plataformas digitales incluye la obligación de evaluar el impacto de las decisiones algorítmicas. En concreto, su artículo 10 establece la obligación de, cada dos años y con la participación de la representación legal de la plantilla, realizar "*una evaluación de los efectos de cada una de las decisiones adoptadas o respaldadas por los sistemas automatizados de supervisión y de toma de decisiones que utilice la plataforma digital de trabajo, para las personas que realizan trabajo en plataformas, en particular, cuando proceda, para sus condiciones laborales y la igualdad de trato en el trabajo*". El uso de sistemas de inteligencia artificial en el ámbito laboral se considera, según establece el artículo 6 en relación con el Anexo III del Reglamento de IA, de alto riesgo y, en consecuencia, se justifica una regulación más proteccionista de los derechos fundamentales de las personas trabajadoras.

## 6. Reflexiones finales

La norma laboral es clara en prohibir toda forma de discriminación en el acceso al empleo; así lo establece de forma clara el artículo 4.2.c) ET. Sin embargo, la discriminación en procesos de selección siempre ha resultado difícil de abordar, al carecerse de la intuición y/o información necesaria para apreciar la existencia de discriminación. La falta de sensación de haber sido discriminado se agudiza, además, con la introducción de sistemas de inteligencia artificial, que se promocionan como modelos matemáticos neutros, objetivos y despojados de toda subjetividad humana.

La discriminación algorítmica, como se ha argumentado en el presente artículo, podrá encauzarse en la actual tutela discriminatoria, constituyendo en muchos supuestos un caso de discriminación indirecta por tratarse de una práctica o tratamiento aparentemente neutro que genera una desventaja particular sobre un colectivo protegido.

[129] West, Sarah Myers, Whittaker, Meredith y Crawford, Kate, "Discriminating Systems: Gender, Race and Power in AI", *AI Now Institute*, 2019, p. 4 (https://ainowinstitute.org/discriminatingsystems.pdf).

Sin embargo, el principal reto para el tratamiento jurídico de la discriminación algorítmica es la opacidad que envuelve a los algoritmos y sistemas de inteligencia artificial, que impide disponer de la información necesaria para apreciar la existencia de discriminación. La actual normativa reconoce derechos de información, pero resultan insuficientes por cuanto no reconocen el acceso a la información necesaria y esencial como es información estadística sobre los efectos del algoritmo.

La transparencia en el uso de sistemas algorítmicos en el ámbito laboral es esencial para evaluar la legalidad de los sistemas de inteligencia artificial y su potencial impacto discriminatorio. En este contexto, mi propuesta es la adopción de una normativa específica referente a la utilización de sistemas algorítmicos y de inteligencia artificial en el ámbito laboral que, además de concretar y ampliar las actuales obligaciones de información, reconozca la obligación de realizar auditorías periódicas sobre el impacto de dichos sistemas y el derecho de la representación legal de la plantilla a acceder a dicha información.

## 7. Bibliografía


Adams-Prassl, Jeremias, "What if your boss was an algorithm? Economic Incentives, Legal Challenges, and the Rise of Artificial Intelligence at Work", *Comparative Labor Law and Policy Journal*, vol. 41, nº 1, 2019, p. 1-30 (versión electrónica).

Adams-Prassl, Jeremias, "When Your Boss Comes Homes", *C4E The Future of Work in the Age of Automation and AI*, 2020, p. 1-11 (https://c4ejournal.net/2020/07/05/jeremias-adams-prassl-when-your-boss-comes-home-2020-c4ej-xxxx-symposium/).

Adams-Prassl, Jeremias, Binns, Reuben y Kelly-Lyth, Aislinn, "Directly Discriminatory Algorithms", *The Modern Law Review*, vol. 86, nº 1, p. 144-175.

Agencia Española de Protección de Datos, "Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción", 2020.

Ajunwa, Ifeoma, Crawford, Kate y Schultz, Jason, "Limitless Worker Surveillance", *California Law Review*, vol. 105, nº 3, 2017, p. 735-776.

Andrews, Lori y Bucher, Hannah, "Automating Discrimination: AI hiring practices and gender inequality", *Cardozo Law Review*, vol 44, nº 1, 2022, p. 145-202.

Article 29 Data Protection Working Party, "Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679", 3.10.2017.

Avogaro, Matteo, "La dirección algorítmica en la propuesta de Directiva sobre el trabajo en plataformas: un avance parcial entre la dimensión individual y colectiva", en Ginès i Fabrellas, Anna (Directora), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 231-265.

Benjamin, Ruha, *Race after technology*, Polity Press, Medford (EUA), 2019.

Bergvall-Kåreborn, Birgitta y Howcroft, Debra, "Amazon Mechanical Turk and the commodification of labor", New Technology, Work and Employment, vol. 29, nº 3, 2014, p. 213-223.

Bolukbasi, Tolga, Chang, Kai-Wei, Zou, James, Saligrama, Venkatesh y Kalai, Adam, "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings", *NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, p. 1-9.

Buolamwini, Joy y Gebru, Timnit, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", *Conference of Fairness, Accountability, and Transparency. Proceedings of Machine Learning Research*, vol. 81, 2018, p. 1-15.

Castillo, Carlos, "Discriminación algorítmica. Aproximación conceptual", en Rivas Vallejo, Pilar (Directora), *Discriminación algorítmica en el ámbito laboral: perspectiva de género e intervención*, Thomson Reuters Aranzadi, 2022, p. 71-77.

Chen, Vickie, "Leveling Up the Gaming Gender Field", *Forbes*, 24.8.2023 (https://www.forbes.com/sites/forbesbusinesscouncil/2023/08/24/leveling-up-the-gaming-gender-gap/).

Cheong, Marc, Lederman, Reeva, Mcloughney, Aidan, Njoto, Sheila, Ruppanner, Leah y Wirth, Anthony, "Gender Occupational Sorting: The role of Artificial Intelligence in Exacerbating Human Bias in STEM Employment", *CIS & Policy Lab*, The University of Melbourne, 26.6.2020, p. 1-12 (version digital).

Cheong, Marc, Lederman, Reeva, Mcloughney, Aidan, Njoto, Sheila, Ruppanner, Leah y Wirth, Anthony, *Ethical implications of AI bias as a result of workforce gender imbalance*. Universidad de Melbourne, 2020.

Choudary, Sangeet Paul, "The architecture of digital labour platforms: policy recommendations on platform design for worker well-being", *ILO Future of Work Research Paper Series*, nº 3, 2018, p. 1-49.

Costa, Allan, Cheung, Chris y Langenkamp, Max, "Hiring Fairly in the Age of Algorithms", *Research Paper Human-Computer Interaction*, Cornell University, 2020, p. 1-33 (version digital).

Crawford, Kate, "The Hidden Biases in Big Data", *Harvard Business Review*, 1.4.2013 (https://hbr.org/2013/04/the-hidden-biases-in-big-data).

Criado Perez, Caroline, Invisible Women. Exposing data bias in a world designed for men, Vintage, Londres, 2019.

Dastin, Jeffrey, "Amazon scraps secret AI recruiting tool that showed bias against women", *Reuters*, 11.10.2018 (https://www.reuters.com/article/idUSKCN1MK0AG/).

Dattner, Ben, Chamorro-Premuzic, Tomas, Buchband, Richard y Schettler, Lucinda, "The Legal and Ethical Implications of Using AI in Hiring", *Harvard Business Review*, 25.4.2019.

de Laat, Paul B., "Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?", *Philosophy & Technology*, nº 31, 2017, p. 1-16 (versión digital).

de Stefano, Valerio, "Algorithmic Bosses and How to Tame Them", *C4E The Future of Work in the Age of Automation and AI*, 2020, p. 1-22 (https://c4ejournal.net/2020/07/05/valerio-de-stefano-algorithmic-bosses-and-how-to-tame-them-2020-c4ej-xxx/).

de Stefano, Valerio, "The rise of the «just-in-time workforce»: On-demand work, crowdwork and labour protection in the «gig-economy»", Conditions of Work and Employment Series, nº 71, Organización Internacional del Trabajo, 2016.

Deva, Surya, "Addressing the gender bias in artificial intelligence and Automation", *Open Global Rights*, 10.4.2020 (https://www.openglobalrights.org/addressing-gender-bias-in-artificial-intelligence-and-automation/).

Dzieza, Josh, "How hard will the robots make us work?", *The Verge*, 27.2.2020 (https://www.theverge.com/2020/2/27/21155254/automation-robots-unemployment-jobs-vs-human-google-amazon).

Eubanks, Virginia, Automating Inequality. How high-tech tools profile, police, and punish the poor, Picador, Nueva York, 2019.

Feldman, Lisa, Adolphs, Ralph, Marsella, Stacy, Martinez, Aleix M. y Pollak, Seth D., "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements", *Psychological Science in the Public Interest*, vol. 20, nº 1, 2019, p. 1-68.

Fernández García, Antonio, "Trabajo, algoritmos y discriminación", en Rodríguez-Piñero Royo, Miguel y Todolí Signes, Adrián (Directores), *Vigilancia y control en el Derecho del Trabajo Digital*, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2020, p. 505-531.

Garben, Sacha, *Protecting workers in the online platform economy: an overview of regulatory and policy developments in the EU*, European Risk Observatory Discussion paper, European Agency for Safety and Health at Work, Luxemburgo, 2017.

Ginès i Fabrellas, Anna, "Algoritmos sesgados en el trabajo. Consideraciones entorno a su tratamiento jurídico", *Trabajo y Derecho*, nº 19, 2024, p. 1-37.

Ginès i Fabrellas, Anna, "Decisiones automatizadas y elaboración de perfiles en el ámbito laboral y su potencial impacto discriminatorio", en Ginès i Fabrellas, Anna (Directora), *Algoritmos, Inteligencia Artificial y Relación Laboral*, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2023, p. 173-229.

Ginès i Fabrellas, Anna, "Disposición addicional 23. Presunción de laboralidad en el ámbito de las plataformas digitales de reparto", en del Rey Gunater, Salvador (Director), *Estatuto de los Trabajadores. Comentado y con jurisprudencia,* La Ley, 4ª edició, Madrid, 2022, p. 2027-2038.

Ginès i Fabrellas, Anna, "El derecho a conocer el algoritmo: una oportunidad perdida de la "Ley Rider", *IUSLabor*, nº 3, 2021, p. 1-5.

Ginès i Fabrellas, Anna, "Sesgos discriminatorios en la automatización de decisiones en el ámbito laboral: evidencias de la práctica", en Rivas Vallejo, Pilar (Directora), *Discriminación algorítmica en el ámbito laboral: perspectiva de género e intervención*, Thomson Reuters Aranzadi, 2022, p. 295-331.

Ginès i Fabrellas, Anna, *El trabajo en plataformas digitales. Nuevas formas de precariedad laboral*, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2021.

Hamilton, Isobel Asher y Cain, Áine, "Amazon warehouse employees speak out about the "brutal" reality of working during the holidays, when 60-hour weeks are man-

datory and ambulance calls are common", *Insider*, 19.2.2019 (https://www.businessinsider.com/amazon-employees-describe-peak-2019-2).

Harwell, Drew, "A face-scanning algorithm increasingly decides whether you deserve the job", *The Washington Post*, 6.11.2019 (disponible en: https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/).

Huws, Ursula, Spencer, Neil H., Syrdal, Dag S. y Holts, Kaire, *Work in the European Gig Economy. Research results from the UK, Sweden, Germany, Austria, The Netherlands, Switzerland and Italy*, Foundation for European Progressive Studies, UNI Europa y University of Hertfordshire, 2017.

Jankowicz, Nina, "Online Harassment Towards Women Is Getting Even More Insidious", *Wired*, 28.1.2021 (https://www.wired.com/story/online-harassment-toward-women-getting-more-insidious/).

Kim, Pauline, "Big Data and Artificial Intelligence: New Challenges for Workplace Equality", *University of Louisville Law Review*, vol. 57, 2019, p. 313-328.

Kosinski, Michal, Stillwell, David y Graepel, Thore, "Private traits and attributes are predictable from digital records of human behavior", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 110, nº 15, 2013, p. 5802-5805.

Kulkarni, Swatee y Che, Xiangdong, "Intelligent Software Tools for Recruiting", *Journal of International Technology and Information Management*, vol 28, nº 2, 2019, p. 1-16.

Kullmann, Miriam, "Platform Work, Algorithmic Decision-Making and EU Gender Equality Law", *International Journal of Comparative Labour Law and Industrial Relations*, vol. 34, nº 1, 2018, p. 1-16 (versión digital).

Kuncel, Nathan R., Ones, Deniz S. y Klieger, David M., "In Hiring, Algorithms Beat Instinct", *Harvard Business Review*, Mayo 2014 (https://hbr.org/2014/05/in-hiring-algorithms-beat-instinct).

Lecher, Colin, "How Amazon automatically tracks and fires warehouse workers for "productivity"", *The Verge*, 25.4.2019 (disponible en: https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations).

Liao, Shannon, "Amazon warehouse workers skip bathroom breaks to keep their jobs, says report", *The Verge*, 16.4.2018 (https://www.theverge.com/2018/4/16/17243026/amazon-warehouse-jobs-worker-conditions-bathroom-breaks).

Lorenz, Taylor y Browning, Kellen, "Dozens of Women in Gaming Speak Out About Sexism and Harassment", *The New York Time*, 23.6.2020 (disponible en: https://www.nytimes.com/2020/06/23/style/women-gaming-streaming-harassment-sexism-twitch.html).

Luque Parra, Manel, "IA y seguridad y salud laboral: la dicotomía entre ser un gran aliado productivo y un "riesgo laboral emergente"", en Ginès i Fabrellas, Anna (Directora), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 305-334.

Malgieri, Gianclaudio y Comandé, Giovanni, "Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation", *International Data Privacy Law*, vol. 7, nº 4, 2017, p. 1-36 (versión electrónica).

Mcfarland, Daniel y Mcfarland, H. Richard, "Big Data and the danger of being precisely inaccurate", *Big Data & Society*, 2015, p. 1-4.

Melchers, Klaus G. y Basch, Johannes M., "Fair play? Sex-, age-, and job-related correlates of performance in a computer-based simulation game", *International Journal of Selection and Assessment*, nº 30, p. 48-61.

Mercader Uguina, Jesús R., *Algoritmos e inteligencia artificial en el derecho del trabajo*, Tirant lo Blanch, Valencia, 2022.

Mercer, *Global Talent Trends 2020*, 2020 (disponible en: https://www.mercer.com/content/dam/mercer/attachments/private/global-talent-trends-2020-report.pdf).

Ministerio de Trabajo y Economía Social, "Información algorítmica en el ámbito laboral. Guía práctica y herramienta sobre la obligación empresarial de información sobre el uso de algoritmos en el ámbito laboral", mayo 2022 (disponible en: https://www.mites.gob.es/ficheros/ministerio/inicio_destacados/Guia_Algoritmos_ES.pdf).

Moore, Phoebe V., "Making Algorithmic Management safe for workers: new regulation is needed", 28.7.2023 (https://phoebevmoore.wordpress.com/2023/07/29/making-algorithmic-management-safe-for-workers-new-regulation-is-needed/?s=09).

Moreno Cáliz, Susana, "Análisis del comportamiento de las plataformas de captación, selección y contratación de trabajadores que emplean algoritmos para la adopción de decisiones: evidencias", en Rivas Vallejo, Pilar (Directora), *Discriminación algorítmica en el ámbito laboral: perspectiva de género e intervención*, Thomson Reuters Aranzadi, 2022, p. 211-232.

Muñoz Ruiz, Ana Belén, Biometría y sistemas automatizados de reconocimiento de emociones: implicaciones jurídico-laborales, Tirant lo Blanch, Valencia, 2023.

Nicks, Leonie, Gesiarz, Filip, Valencia, Lourdes, Hardy, Tim y Lohmann, Johannes, *Gender differences in response to requirements in job adverts*, The Behavioural insights team, Government Equalities Office, 2022.

O'Neil, Cathy, Weapons of Math Destruction. How Big Data increases inequality and threatens democracy, Penguin Books, Reino Unido, 2016.

Pastor Martínez, Alberto, "Los derechos colectivos de información, consulta y negociación del uso de algoritmos y sistemas de Inteligencia Artificial", en Ginès i Fabrellas, Anna (Directora), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 335-361.

Peck, Don, "They're Watching You at Work", *The Atlantic*, diciembre 2013 (https://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/).

Pérez del Prado, Daniel, Derecho, economía y digitalización. El impacto de la inteligencia artificial, los algoritmos y la robótica sobre el empleo y las condiciones de trabajo, Tirant lo Blanch, Valencia, 2023.

Peters, Jay, "Internal documents show automated Amazon warehouses have higher injury rates", *The Verge*, 29.9.2020 (https://www.theverge.com/2020/9/29/21493752/amazon-warehouses-robots-higher-injury-rates-report-reveal).

Prassl, Jeremias, *Humans as a service. The promise and perils of work in the gig economy*, Oxford University Press, Nueva York, 2018.

Preciado Domenech, Carlos Hugo, "Algoritmos y discriminación en la relación laboral", *Jurisdicción Social*, nº 223, 2021, p. 5-24.

Raub, McKenzie, "Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices", *Arkansas Law Review*, vol. 71, nº 2, 2018, p. 529-570.

Rivas Vallejo, Pilar, La aplicación de la Inteligencia Artificial al trabajo y su impacto discriminatorio, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2020.

Rogers, Brishen, "The Social Costs of Uber", The University of Chicago Law Review Dialogue, vol. 82, 2015, p. 85-102.

Rosenblat, Alex, *Uberland. How algorithms are rewriting the rules of work*, University of California Press, Oakland (Estados Unidos), 2018.

Rudin, Cynthia, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisiones and Use Interpretable Models Instead", *Nature Machine Intelligence*, vol. 1, 2019, p. 1-20 (versión digital).

Sáez Lara, Carmen, "El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación", *Temas Laborales*, nº 155, 2020, p. 41-60.

Schellmann, Hilke, "Auditors are testing hiring algorithms for bias, but there's no easy fix", *MIT Technology Review*, 11.2.2021.

Schellmann, Hilke, *The Algorithm. How AI decides who gets hired, monitored, promoted & fired & why we need to fight back now*, Hachette Books, Nova York, 2024.

Todolí Signes, Adrián, Algoritmos productivos y extractivos. Cómo regular la digitalización para mejorar el empleo e incentivar la innovación, Aranzadi, Cizur Menor (Navarra), 2023.

Unceta, Irene, "Notas para un aprendizaje automático justo", en Ginès i Fabrellas, Anna (Directora), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 81-111.

UNESCO, "Women's access to and participation in technological developments" (https://www.unesco.org/en/artificial-intelligence/gender-equality).

UNESCO, Challenging systematic Prejudices: an investigation into Gender Bias in Large Language Models, 2024

Véliz, Carissa, *Privacy is power. Why and how you should take back control of your data*, Transworld publishers, Londres, 2020.

Vincent, James, "Amazon reportedly scraps internal AI recruiting tool that was biased against women", *The Verge*, 10.10.2018 (https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G).

Wachter, Sandra y Mittelstadt, Brent, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI", *Columbia Business Law Review*, vol. 2, 2019, p. 494-620.

Wachter, Sandra, Mittelsdtadt, Brent y Russell, Chris, "Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI", *Computer Law & Security Review*, vol. 41, 2021, p. 1-72 (version digital).

Wall, Sheridan y Schellmann, Hilke, "LinkedIn's job-matching AI was biased. The company's solution? More AI", *MIT Technology Review*, 23.6.2021.

West, Sarah Myers, Whittaker, Meredith y Crawford, Kate, "Discriminating Systems: Gender, Race and Power in AI", *AI Now Institute*, 2019 (https://ainowinstitute.org/discriminatingsystems.pdf).

Zou, James, "Removing gender bias from algorithms", *The Conversation*, 26.9.2019 (https://theconversation.com/removing-gender-bias-from-algorithms-64721).

Zuboff, Shoshana, *The age of surveillance capitalism. The fight for a human future at the new frontier of power*, Profile Books, Reino Unido, 2019.

# Procesos de selección algorítmica y discriminación (I)

## Algorithmic selection and discrimination processes (I)

Gemma Fabregat Monfort
*Catedrática de Derecho del Trabajo y de la Seguridad Social*
*Universitat de València*

doi: 10.20318/labos.2024.9035

*Resumen:* El presente trabajo es la traslación escrita de la conferencia cuya temática compartí con la profesora Ginès i Fabrellas en el Congreso denominado La IA en el mundo del trabajo que se desarrolló en la Facultad de Derecho de la Universitat de València bajo la brillante dirección de los profesores Todolí y Beltrán.
Su objetivo es analizar las posibles vulneraciones que a la prohibición de discriminación pueden generarse en el procedimiento de selección algorítmica de las personas trabajadoras. Tras fijar el marco normativo aplicable, ciertamente disperso, se determinan los pormenores de estas situaciones discriminatorias planteando, especialmente, si ante las dificultades que genera su detección, el actual marco legal posibilita su tutela efectiva.
*Palabras clave:* Selección algorítmica, prohibición de discriminación, tipos de discriminación, tutela legal.

*Abstract:* This paper is the written translation of the conference whose theme I shared with Professor Ginès i Fabrellas at the Congress entitled 'AI in the world of work', which was held at the Faculty of Law of the University of Valencia under the brilliant direction of Professors Todolí and Beltrán.
Its objective is to analyse the possible violations of the prohibition of discrimination that may arise in the procedure for algorithmic selection of workers. After establishing the applicable legal framework, which is certainly scattered, the details of these discriminatory situations are determined by considering, in particular, whether, given the difficulties that generate their detection, the current legal framework makes effective protection possible.
*Keywords:* Algorithmic selection, prohibition of discrimination, types of discrimination, legal protection.

## Introducción

El presente artículo recoge las ideas centrales expuestas en la conferencia que, con el mismo título, tuve la posibilidad de impartir junto con la profesora Ginès i Fabrellas en el Congreso denominado *La IA en el mundo del trabajo* que se desarrolló en la Facultad de

Derecho de la Universitat de València bajo la brillante dirección de los profesores Todolí y Beltrán.

Los profesores Goerlich y Mercader, ponentes también en el Congreso, han tenido a bien poner a nuestra disposición una edición especial de la revista que dirigen junto con la profesora De la Puebla, dedicando un monográfico al tan importante como novedoso tema de la IA en el mundo laboral.

Cuanto sigue, así, es el resultado de la investigación o estudio realizado *ad hoc* para el mencionado Congreso. La temática, como decía, fue compartida con la profesora Ginès. Nuestras intervenciones, tras coordinarnos, creo que se complementaron sin solaparse. Me ciño pues, en cuanto sigue, al mismo esquema que seguí en la intervención oral. La pretensión no es tanto ser fiel a lo que en su día expuse, que también, como no reiterar en el tratamiento escrito de la ponencia posibles cuestiones abordables en el artículo que por su parte firma la profesora Ginès i Fabrellas.

## 1. La gestión algorítmica en el acceso a la empresa

Una de las cuestiones que más preocupan en el ámbito jurídico laboral de un tiempo a esta parte es la del impacto de la IA en el mundo del trabajo. Especialmente, aunque no exclusivamente, dada su efectividad estadística, en lo referente a los procedimientos de gestión de las personas y, entre estos, a los de selección de personal[1].

[1] Al respecto, véase, por todos, *Big data, algorithms and discrimination,* 2018, https://fra.europa.eu/sites/default/files/fra-2018-in-brief-bigdata-algorithms-disrimination, pdf; Recomendación CM/Rec(2020)1 de 8 de abril de 2020, del Comité de Ministros del Consejo de Europa, sobre impacto en los derechos humanos de los sistemas algorítmicos, disponible en https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154 . También de mucho interés es la Recomendación CM/Rec(2021)8, sobre protección de las personas en relación con el tratamiento automatizado de datos en el contexto de la elaboración de perfiles, disponible en https://search.coe.int/cm/pages/result_details.aspx?ObjectId=0900001680a46147.

Por parte de la doctrina, entre otros, GINES I FABRELLAS, A., "Sesgos discriminatorios en la automatización de decisiones en el ámbito laboral: evidencias de la práctica". En RIVAS VALLEJO, P., *Discriminación algorítmica en el ámbito laboral: perspectiva de género e intervención*. Aranzadi. Pamplona, 2022; GINÉS I FABRELLAS, A., *Algoritmos, inteligencia artificial y relación laboral*. Editorial Aranzadi. Pamplona, 2023; OLARTE ENCABO, S., "La aplicación de inteligencia artificial a los procesos de selección de personal y ofertas de empleo. Impacto sobre el derecho a la no discriminación". Documentación Laboral, n.º 119, 2020, págs. 95-97; MERCADER UGUINA, J. R., "Discriminación algorítmica y derecho granular: nuevos retos para la igualdad en la era del Big data". LABOS Revista De Derecho Del Trabajo y Protección Social, n.º 2, 2021, págs. 4-10; RODRIGUEZ CARDO, I., "Decisiones automatizadas y discriminación algorítmica en la relación laboral ¿hacia un Derecho del Trabajo de dos velocidades?". Revista Española de Derecho del Trabajo, n.º 253, 2022, págs. 135-188; PEYRONNET, M., "El uso de los algoritmos y la inteligencia artificial en la selección de personal: ¿una ocasión para eliminar la discriminación". En MOLINA NAVARRETE, C., y VALLECILLO GÁMEZ, M. R. (Dir.), *De la economía digital a la sociedad del E-Work decente: condiciones sociolaborales para una industria 4.0 justa e inclusiva.* Aranzadi. Pamplona, 2021; VIOLLIER, P., y VELASCO, P., "El uso de toma de decisiones automatizadas para la selección de personal". En SEVERÍN CONCHA, J.P. (Ed.), *Derechos fundamentales de la persona del trabajador.* Tirant lo Blanch. Valencia, 2021.

Sin duda no puede negarse que efectivamente la IA y ciertos algoritmos[2] pueden ser muy útiles en el proceso de *cribado* que puede exigir en determinados casos elegir quien, entre las personas candidatas a ocupar un puesto de trabajo, parece ser la persona que mejor encajaría en el perfil buscado por la empresa.

Con todo, la eficacia o la efectividad de la IA a estos efectos, poco discutible fundamentalmente en determinados sectores y puestos de trabajo, debe coordinarse con las garantías constitucionales que para con las personas trabajadoras reconoce nuestro ordenamiento jurídico, especialmente en todo lo que tiene que ver con la tutela de los derechos fundamentales y la prohibición de la discriminación.

Para empezar, hay que resaltar un hecho tan evidente como en ocasiones silenciado, cual es, que la selección algorítmica es una decisión empresarial[3].

Es la empresa la que decide o no instrumentalizar el proceso de selección recurriendo a la IA o al algoritmo. Pero la decisión de contratar o no, o de seleccionar a una persona y no a otra, no es del algoritmo, es de la empresa. Aunque ésta la adopte siguiendo las pautas, las indicaciones o los resultados de un algoritmo o de un sistema de IA más complejo[4].

Lo anterior debe ser el punto de partida. Pues en la medida en que, por mucho que se recurra a un algoritmo, se tenga claro que la decisión de seleccionar, o no, es en realidad una decisión que forma parte de la potestad empresarial, los límites a aplicar también lo serán.

Conviene subrayar que en nuestro ordenamiento jurídico se reconoce en el art. 38 CE la libertad de empresa dentro de una economía de mercado. De esa libertad de em-

[2] En definición de la RAE el algoritmo es "el conjunto ordenado y finito de operaciones que permite hallar la solución de un problema"; y la inteligencia artificial, por su parte, es definida como la "Disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico".

[3] Y ello aun y cuando la decisión se adopte de manera prácticamente automatizada, esto es, en procedimientos con poca o nula intervención humana. En cualquier caso, y con independencia de lo anterior y del control humano o humano al mando hoy expresamente exigido por el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo 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 y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) -en adelante, RGPD- y 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.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). –en adelante, RIA–, en los procesos de selección en los que intervenga la IA, lo cierto es que en el ámbito de la relación laboral el que responde ante el candidato es el empleador. El algoritmo es un instrumento del empleador, a su disposición para hacer más eficiente, en ocasiones, la selección de las personas a contratar. Pero el que responde ante el derecho es el empleador, el sujeto, no el algoritmo, que es el instrumento del que se vale este para, como digo, seguramente ser más eficaz eligiendo a quien contratar.

[4] El algoritmo no tiene personalidad jurídica, por lo que difícilmente puede ser titular de derechos y obligaciones. La decisión es, pues empresarial, aunque se base en la opción seleccionada por el algoritmo. Cuestión distinta, como se verá, es que se prevean una serie de garantías para evitar daños en las decisiones automatizadas, entendiendo como tales aquellas que acata el empresario siguiendo las sugerencias del algoritmo sin básicamente cuestionarlas o reconsiderarlas.

presa emanan, entre otras, la decisión empresarial de contratar a quien se desee contratar, la decisión de no hacerlo, o la decisión de contratar a una persona y no a otra. Cualquiera de estas opciones forma parte de la discrecionalidad empresarial. Discrecionalidad, que no arbitrariedad, ante la existencia de unos límites muy concretos: el respeto a los derechos fundamentales y la prohibición de discriminación de las personas implicadas.

La ubicación física y numérica en el texto constitucional de los mencionados derechos evidencia, por sí misma, la relevancia de lo primero respecto lo segundo. O lo que es igual, el obligado sometimiento de cualquier decisión adoptada en base al art. 38 CE, a los mandatos constitucionales que lo preceden, tal y como desde hace tiempo ya desarrolla la conocida como teoría de la triple proporcionalidad y que solamente acepta la afectación sin vulneración de un derecho fundamental por parte de la empresa en los casos en los que la decisión resulte ser idónea, necesaria y proporcional entre el agravio individual y el beneficio de la colectividad[5].

Nada de eso se da de manera *apriorística* en un proceso de selección. Con o sin algoritmo, no se puede acometer un proceso de selección vulnerando derechos fundamentales de las personas candidatas tales como pueden ser, por ejemplo, la intimidad o la dignidad de la persona candidata, ni tampoco valorar a esos efectos determinadas causas discriminatorias. Solo excepcionalmente, y de concurrir justificación objetiva en función del puesto a ocupar, se podría tener en cuenta, el sexo de las personas a contratar, su estado de salud o cualquier otra de las causas susceptibles de generar discriminación. Lo primero, ocurrirá, por ejemplo, para cumplir con una medida de acción positiva o en aplicación de los porcentajes de Ley Orgánica 2/2024, de 1 de agosto, de representación paritaria y presencia equilibrada de mujeres y hombres[6] a la que después me referiré.

El sexo de las personas que concurren a un proceso de selección solamente será un elemento a ponderar, cuando (por decirlo en palabras de la ley[7]) *debido a la naturaleza de las actividades profesionales concretas o al contexto en el que se lleven a cabo, dicha característica constituya un requisito profesional esencial y determinante, siempre y cuando el objetivo sea legítimo y el requisito proporcionado*. Lo que, de forma más extensiva para cualquier otra causa susceptible de generar discriminación, se reproduce en lo esencial en el art. 4.2[8] Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, con especial mención al estado de salud en el art. 2.3. de la misma Ley 15/2022, en el sentido de que *la enfermedad no podrá amparar diferencias de trato distintas de las que deriven del propio proceso de tratamiento de la misma, de las limitaciones*

[5] La STC 99/1994, de 11 de abril fue la primera que en el ámbito laboral instó a aplicar el mencionado triple juicio de proporcionalidad a los conflictos entre poder de dirección y derechos fundamentales de las personas trabajadoras.

[6] En adelante, Ley Orgánica 2/2024, de Paridad.

[7] En aplicación del art. 5, segundo inciso, de la Ley 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres –en adelante, LOI–.

[8] No se considera discriminación la diferencia de trato basada en alguna de las causas previstas en el apartado 1 del artículo 2 de esta ley derivada de una disposición, conducta, acto, criterio o práctica que pueda justificarse objetivamente por una finalidad legítima y como medio adecuado, necesario y proporcionado para alcanzarla.

*objetivas que imponga para el ejercicio de determinadas actividades o de las exigidas por razones de salud pública.* Teniendo en cuenta, además, que solamente en el caso de que el puesto de trabajo a desarrollar se incluyese entre los posibles a generar una enfermedad profesional, en aplicación del art. 243.1 LGSS, el estado de salud de la persona candidata podrá ser objeto de vigilancia empresarial de forma previa a la formalización de un contrato de trabajo[9].

El problema, en el caso de la IA, es que si bien aclarado lo anterior, el mandato de la regulación normativa resulta bastante evidente, y ni la empresa puede *ocultarse* detrás de un algoritmo para obviar el mandato constitucional, ni a la inversa, el volumen de datos, la opacidad del algoritmo, y la eficiencia como finalidad sin la eliminación de sesgos, dificultan considerablemente confrontar el posible cumplimiento o incumplimiento empresarial en el sentido expuesto, fundamentalmente de forma previa a la formalización de la contratación[10].

Producida ésta, desde la perspectiva de la discriminación indirecta, y con el análisis del impacto de los resultados alcanzados, se podrán detectar posibles discriminaciones. Pero de manera previa a la materialización de la selección, a la decisión de seleccionar, no siempre será sencillo identificar la concurrencia de discriminación. La discriminación es un perjuicio por prejuicio(s). Y los mismos prejuicios o sesgos conscientes o inconscientes que pueden influir en la selección u elección de una persona y no de otra en una selección previa a la contratación sin presencia de IA, pueden estar en la mente de quienes han participado en el diseño y elaboración del algoritmo determinante de la selección[11] y/o en la selección de los datos de los que se *alimenta* el algoritmo para tomar una de-

[9] Art. 243 LGSS.1. Todas las empresas que hayan de cubrir puestos de trabajo con riesgo de enfermedades profesionales están obligadas a practicar un reconocimiento médico previo a la admisión de los trabajadores que hayan de ocupar aquellos y a realizar los reconocimientos periódicos que para cada tipo de enfermedad se establezcan en las normas que, al efecto, apruebe el Ministerio de Empleo y Seguridad Social.

[10] "Riesgos que aumentan exponencialmente cuando se da el salto del mero tratamiento de datos al empleo de la inteligencia artificial, ya que ésta comporta la reproducción, mediante una máquina de habilidades cognitivas humanas, generalmente aplicando algoritmos que se basan en modelos diseñados a través de aprendizaje automático a partir de la entrada de información, en muchos casos, de datos personales, no siempre estrictamente profesionales. Aparentemente, la aplicación de IA a los procesos de intermediación laboral y de selección de personal, por su objetividad y asepsia, podría dar lugar a un funcionamiento más eficiente del mercado de trabajo, blindado, además, frente a cualquier forma de discriminación, pero la realidad es bien distinta, ya que no se controla ni se conocen los datos que se suministran, ni de donde se obtienen, ni los perfiles que interesan o se descartan, con el resultado de una exclusión apriorística de personas con las competencias profesionales requeridas; tampoco se conocen las instrucciones que va automatizando la máquina, es decir, los inputs y data input. Por tanto, el avance digital plantea nuevos desafíos jurídicos en la efectividad y tutela del derecho fundamental de igualdad y no discriminación en el acceso al empleo ya que, por su complejidad técnico-informática, las prácticas discriminatorias se hacen opacas e inextricables, estando además su soporte técnico –generalmente algoritmos– amparado por el secreto industrial (y esto sí constituye una novedad)" (OLARTE ENCABO, S., "La aplicación de inteligencia…", cit. pág. 82).

[11] "La evaluación de candidatos mediante inteligencia artificial se articula mediante la priorización de características que se asocian a determinados rasgos o competencias que el diseñador del algoritmo introduce según instrucciones del empleador ofertante de puestos de trabajo (…)" (OLARTE ENCABO, S., "La aplicación de inteligencia…", cit. pág. 82).

cisión y no otra[12]. Y, como decía, el inabarcable volumen de datos de los que se nutren y con los que se entrenan los algoritmos y la opacidad de la fórmula matemática que lo conforma, dificultan sustancialmente el control de la no discriminación en su aplicación. Fundamentalmente si el control se quiere ejercer de forma previa a la materialización de la selección, antes de la conclusión del proceso selectivo[13].

Además, y por si lo anterior fuese poco, que sin duda no lo es, la discriminación no siempre es abstracta. La posible concurrencia de motivos objetivos que excepcionalmente pueden permitir en atención al puesto de trabajo a ocupar, considerar causas en otros casos no permitidas, en otras ocasiones exigiría ajustar los algoritmos a cada empresa y, dentro de esta, a cada proceso selectivo en cuestión, considerando singularmente premisas no generalistas[14].

Y esto es todavía más difícil cuando la selección se hace descansar sobre algoritmos previamente diseñados por un tercero externo a la empresa, que lo adquiere y lo aplica (habitualmente) sin los matices que podría suponer tener en cuenta en cada caso en concreto determinadas circunstancias que, si bien en unos casos pueden ser razonables, por no responder a justificación objetiva, en otros procedimientos selectivos podrían resultar constitutivos de discriminación precisamente por lo contrario, esto es, por no concurrir en ellas causa objetiva de justificación.

---

[12] "Sin embargo, algoritmo, sesgo y discriminación son términos que se conectan frecuentemente. En efecto, un aspecto crítico de los sistemas de IA es el de la posible existencia de sesgos o "bias", que serían desviaciones inadecuadas en el proceso de inferencia. Aunque el sesgo de la decisión no es un problema particular de los sistemas de IA, sino que es general de cualquier proceso de toma de decisión, ya sea humano o automático, si es preocupante que la elaboración de perfiles pueda perpetuar los estereotipos existentes y ampliar la segregación social" (SAEZ LARA, C., "El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación". Temas laborales: Revista andaluza de trabajo y bienestar social, n.º 155, 2020, págs. 42-60).

"Hay numerosas pruebas de que los algoritmos pueden cobijar sesgos ocultos. Por un lado, se corre el riesgo de que, sin pretenderlo deliberadamente, acaben reflejando los de sus programadores. Y, por otro lado, el entrenamiento de datos puede llevarse a cabo con información sesgada. Es lo que se sintetiza con la expresión: 'entra basura, sale basura' (o 'garbage in, garbage out' – GIGO) (BELTRÁN DE HEREDIA RUIZ, I., "Algoritmos psicometría y derechos del yo inconsciente de la persona en el ámbito socio-laboral". Revista Derecho Social y Empresa, n.º 18, 2023 (Ejemplar dedicado a: La tecnología y la digitalización en las relaciones laborales: personas y competitividad), págs. 154-179".

[13] Es de sobra conocido el caso del software de Amazon que se entrenó solo con curriculums masculinos, lo que le llevaba a excluir a las mujeres con formación técnica adecuada para puestos técnicos, por considerar el algoritmo que ser varón era requisito determinante para poder ejercer de técnico. Tras intentos por parte de Amazon de modificar el sesgo, tres años más tarde de su creación, eliminó el programa de reclutamiento.

En todo caso, es importante destacar que en realidad lo que discriminaba no era el algoritmo, o no solo. El algoritmo, paradójicamente, aquí permitió evidenciar una práctica que durante mucho tiempo había seguido Amazon, cual era, la de no contratar a mujeres para puestos técnicos.

[14] No podemos ignorar, además, que si bien se ha interpretado de forma restrictiva, los datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o las orientación sexuales de una persona física, deben tener en cuanto datos personales un tratamiento especial según el art. 9 del RGPD.

Sin olvidar, por lo demás, que los mismos sistemas de IA han evolucionado tanto que son capaces de aprender de sí mismos y alcanzar deducciones o realizar sugerencias presuntamente basadas en la eficacia en base a parámetros muy distintos de aquellos en sobre los que se concibieron en su diseño o programación.

Veamos, por tanto, qué instrumentos normativos pueden resultar aplicables a esta realidad, con el fin de analizar de qué manera desde la ley se puede contribuir a clarificar cuando un proceso de selección empresarial instrumentalizado mediante el recurso a un algoritmo puede resultar constitutivo de discriminación y de qué forma puede minimizarse el riesgo de que así sea.

## 2. Marco Normativo aplicable

En el elenco de normas aplicables al tema que nos ocupa debe diferenciarse entre normativa supranacional y nacional. Y, en ambos casos, más allá de las genéricas, las específicas tienen que ver bien con la prohibición de discriminación; bien con la regulación de la IA.

### *2.1. Ámbito supranacional*

Desde una perspectiva supranacional, al respecto debe tomarse en consideración fundamentalmente lo siguiente:

— El Convenio núm. 111 OIT, relativo a la discriminación en materia de empleo y ocupación.
— Las Directivas y Reglamentos:

- Directiva 76/207/ CEE, Directiva del Consejo, de 9 de febrero de 1976, relativa a la aplicación del principio de igualdad de trato entre hombres y mujeres en lo que se refiere al acceso al empleo, a la formación y a la promoción profesionales, y a las condiciones de trabajo, modificada parcialmente por la Directiva 2002/73/CE del Parlamento Europeo y del Consejo, de 23 de septiembre de 2002, que modifica la Directiva 76/207/CEE del Consejo relativa a la aplicación del principio de igualdad de trato entre hombres y mujeres en lo que se refiere al acceso al empleo, a la formación y a la promoción profesionales, y a las condiciones de trabajo, y derogada por la Directiva 2006/54/CE del Parlamento Europeo y del Consejo, de 5 de julio de 2006, relativa a la aplicación del principio de igualdad de oportunidades e igualdad de trato entre hombres y mujeres en asuntos de empleo y ocupación (refundición). Así como la Directiva (UE) 2024/1500 – DUE de 29 de mayo-del Parlamento Europeo y del Consejo, de 14 de mayo de 2024, sobre las normas relativas a los organismos de igualdad en el ámbito de la igualdad de trato y la igualdad

de oportunidades entre mujeres y hombres en materia de empleo y ocupación, y por la que se modifican las Directivas 2006/54/CE y 2010/41/UE, y que todavía no ha sido objeto de transposición por nuestro país[15].

- Directiva 79/7/CEE, Directiva del Consejo, de 19 de diciembre de 1978, relativa a la aplicación progresiva del principio de igualdad de trato entre hombres y mujeres en materia de seguridad social; Directiva 2000/43/CE, del Consejo, de 29 de junio de 2000, relativa a la aplicación del principio de igualdad de trato de las personas independientemente de su origen racial o étnico. Su capítulo III, se suprime por Directiva 2024/1499, de 7 de mayo (Ref. DOUE-L-2024-80810). Se transpone por Ley 62/2003, de 30 de diciembre.
- Directiva 2000/78/CE del Consejo, de 27 de noviembre de 2000, relativa al establecimiento de un marco general para la igualdad de trato en el empleo y la ocupación. Se transpone por Ley 62/2003, de 30 de diciembre.
- Directiva del Consejo 2004/113/CE, de 13 de diciembre de 2004, por la que se aplica el principio de igualdad de trato entre hombres y mujeres al acceso a bienes y servicios y su suministro. Se suprime el capítulo III, por Directiva 2024/1499, de 7 de mayo (Ref. DOUE-L-2024-80810). Se anula con efectos de 21 de diciembre de 2012, el art. 5.2, por Sentencia de 1 de marzo de 2011 (Ref. DOUE-Z-2011-70005). Se transpone por la Ley Orgánica 3/2007, de 22 de marzo (Ref. BOE-A-2007-6115).
- 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.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).
- Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo 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 y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos).

### *2.2. Ámbito nacional*

Desde el punto de vista interno, deben considerarse:

— Constitución Española, C.E., especialmente los arts. 9.2, el Título Primero, especialmente arts. 10, 14, 16, 18, 20, 38 CE.

[15] La Directiva, publicada en el Diario Oficial de la Unión Europea el 29 de mayo de 2024, establece una serie de requisitos mínimos que los Estados Miembros deben cumplir en relación con los organismos de igualdad, las entidades responsables de promover y defender la igualdad en el trabajo.

— Estatuto de los Trabajadores, también con especial consideración art. 4.2, 17, etc.
— LOI, y su art. 5, entre otros
— Real Decreto Legislativo 1/2013, de 29 de noviembre, por el que se aprueba el Texto Refundido de la Ley General de derechos de las personas con discapacidad y de su inclusión social.
— Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.
— Ley 22/2021, de 28 de diciembre, de Presupuestos Generales del Estado para el año 2022[16].
— La Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación.
— Ley 4/2023, de 28 de febrero, para la igualdad real y efectiva de las personas trans y para la garantía de los derechos de las personas LGTBI
— Ley Orgánica 2/2024, de 1 de agosto, de representación paritaria y presencia equilibrada de mujeres y hombres.

## 3. Tipología de las conductas discriminatorias

De cuanto conforma este marco normativo, sin duda extenso, me centraré, ahora, en lo relativo a las distintas manifestaciones de la discriminación, conceptos a veces novedosos, con el fin de evidenciar qué actuación de los algoritmos y la IA puede resultar discriminatorias incluso sin querer serlo.

De hecho, lo que pretendo en este epígrafe no es tanto señalar las causas que según nuestro ordenamiento jurídico pueden comportar discriminación, sino las conductas que en sí mismas jurídicamente tienen la consideración de discriminatorias.

Tras marcar claramente en los años 80 la diferenciación entre la igualdad formal y la igualdad real del primer y segundo inciso del art. 14 CE en relación con el art. 9.2. CE[17], y la interrelación de esta última con la prohibición de discriminación, el TC[18] dio un paso más en la tutela de las conductas discriminatorias reconociendo expresamente en nuestro ordenamiento jurídico, junto con el concepto de discriminación directa, el concepto de discriminación indirecta. La distinción entre una y otra manifestación de discriminación desde siempre ha sido relevante. Desde la perspectiva de la discrimi-

[16] Esta norma, aunque a priori no tiene nada que ver ni con la discriminación ni con algoritmos o IA, contempla en su DA 130.ª la creación de una Agencia Española de Supervisión de la Inteligencia Artificial, que deberá minimizar los "riesgos significativos sobre la seguridad y salud de las personas, así como sobre sus derechos fundamentales, que puedan derivarse del uso de sistemas de inteligencia artificial", por eso lo señalo expresamente. La Agencia (AESIA) finalmente inició sus actividades el pasado mes de junio, presentándose, además, al tiempo, a su director general.

[17] STC 83/1984 de 24 de julio.

[18] La primera fue la STC 145/1991, de 1 de julio. Aunque hay algunos autores que afirman que, en realidad, la primera referencia judicial al respecto se contiene en el voto particular del magistrado Rubio Llorente a la STC 103/1983, de 22 de noviembre.

nación indirecta se considera que existe discriminación cuando un criterio o práctica neutra genera un efecto adverso en un colectivo respecto otro sin que exista un motivo objetivo que lo justifique.

La discriminación indirecta es colectiva que no individual (si alguien la sufre es porque forma parte de un colectivo, no por sí mismo); no exige intencionalidad, por cuanto que lo relevante es el efecto adverso y, además, acepta prueba en contra. El efecto adverso en un colectivo no es discriminatorio en cuanto tal, si efectivamente existe una razón objetiva que justifica tal criterio, práctica o prueba.

Estando así las cosas, y siendo hasta la fecha básicamente utilizados por los tribunales los conceptos de discriminación directa e indirecta en materia de discriminación por razón de sexo y de género, la entrada en vigor de la LOI, fijó como mandato legal lo que ya era un consolidado concepto jurisprudencial en el sentido expuesto.

En efecto, con una definición casi idéntica a la que ya jurisprudencialmente se utilizaba, el art. 6 LOI contempla lo siguiente:

> *Artículo 6. Discriminación directa e indirecta.*
>
> *1. Se considera discriminación directa por razón de sexo la situación en que se encuentra una persona que sea, haya sido o pudiera ser tratada, en atención a su sexo, de manera menos favorable que otra en situación comparable.*
>
> *2. Se considera discriminación indirecta por razón de sexo la situación en que una disposición, criterio o práctica aparentemente neutros pone a personas de un sexo en desventaja particular con respecto a personas del otro, salvo que dicha disposición, criterio o práctica puedan justificarse objetivamente en atención a una finalidad legítima y que los medios para alcanzar dicha finalidad sean necesarios y adecuados.*
>
> *3. En cualquier caso, se considera discriminatoria toda orden de discriminar, directa o indirectamente, por razón de sexo.*

Además, esa misma ley especifica en su art. 8 un tipo de discriminación directa por razón de sexo, en concreto, **la discriminación por embarazo o maternidad, en los siguientes términos**:

> *Constituye discriminación directa por razón de sexo todo trato desfavorable a las mujeres relacionado con el embarazo o la maternidad*

Años después, y ya no solo vinculado al sexo, sino a otras muchas causas de discriminación, la Ley 15/2022, vuelve a dar un concepto normativo de discriminación directa e indirecta. Lo que ocurre es que la definición de las distintas manifestaciones de la discriminación en la Ley 15/2022 no se limita a esta doble conducta. Junto con estos dos conceptos, esa nueva Ley Integral de Igualdad expresamente contempla otras manifestaciones de discriminación hasta ahora novedosas en nuestro ordenamiento jurídico interno y que, a los efectos que aquí interesan, deben igualmente tutelarse en el proceso de selección, en el momento del acceso a la empresa. Obviamente, también cuando este se mediatiza recurriendo a un algoritmo.

Al margen del acoso discriminatorio o las represalias, el art. 6 de la Ley 15/2022 define la discriminación directa y la indirecta desde una perspectiva causal más general,

la discriminación por asociación y por error; y la discriminación múltiple e interseccional, advirtiendo que también es discriminatoria la inducción, orden o instrucción de discriminar. Lo hace en los siguientes términos:

1. Discriminación directa e indirecta.

   a) La discriminación directa es la situación en que se encuentra una persona o grupo en que se integra que sea, haya sido o pudiera ser tratada de manera menos favorable que otras en situación análoga o comparable por razón de las causas previstas en el apartado 1 del artículo 2.
   Se considerará discriminación directa la denegación de ajustes razonables a las personas con discapacidad. A tal efecto, se entiende por ajustes razonables las modificaciones y adaptaciones necesarias y adecuadas del ambiente físico, social y actitudinal que no impongan una carga desproporcionada o indebida, cuando se requieran en un caso particular de manera eficaz y práctica, para facilitar la accesibilidad y la participación y garantizar a las personas con discapacidad el goce o ejercicio, en igualdad de condiciones con las demás, de todos los derechos.
   b) La discriminación indirecta se produce cuando una disposición, criterio o práctica aparentemente neutros ocasiona o puede ocasionar a una o varias personas una desventaja particular con respecto a otras por razón de las causas previstas en el apartado 1 del artículo 2.

2. Discriminación por asociación y discriminación por error.

   a) Existe discriminación por asociación cuando una persona o grupo en que se integra, debido a su relación con otra sobre la que concurra alguna de las causas previstas en el apartado primero del artículo 2 de esta ley, es objeto de un trato discriminatorio.
   b) La discriminación por error es aquella que se funda en una apreciación incorrecta acerca de las características de la persona o personas discriminadas.

3. Discriminación múltiple e interseccional.

   a) Se produce discriminación múltiple cuando una persona es discriminada de manera simultánea o consecutiva por dos o más causas de las previstas en esta ley.
   b) Se produce discriminación interseccional cuando concurren o interactúan diversas causas de las previstas en esta ley, generando una forma específica de discriminación.
   c) En supuestos de discriminación múltiple e interseccional la motivación de la diferencia de trato, en los términos del apartado segundo del artículo 4, debe darse en relación con cada uno de los motivos de discriminación.

d) Igualmente, en supuestos de discriminación múltiple e interseccional las medidas de acción positiva contempladas en el apartado 7 de este artículo deberán atender a la concurrencia de las diferentes causas de discriminación.

5. Inducción, orden o instrucción de discriminar.

a) Es discriminatoria toda inducción, orden o instrucción de discriminar por cualquiera de las causas establecidas en esta ley.
b) La inducción ha de ser concreta, directa y eficaz para hacer surgir en otra persona una actuación discriminatoria.

## 4. El acceso a la empresa y la discriminación en conductas omisivas

El marco normativo expuesto, sin duda, supone un avance considerable en el reconocimiento legal de las conductas prohibidas por discriminatorias. Sin embargo, conciliar estos conceptos con un acto omisivo como supone el no contratar o no seleccionar mediante un algoritmo supone, como se decía, cierta complejidad.

Desde siempre, la más difícil demostración de la discriminación se ha dado en las conductas omisivas. En el no hacer, en el no seleccionar, en el no contratar. Demostrar que no se selecciona a una persona en beneficio de otra de forma discriminatoria ha sido siempre uno de los obstáculos de la tutela discriminatoria sancionadora. Solo el concepto de discriminación indirecta aplicado a los resultados de esa acción puede demostrar *a posteriori* que la discriminación ha existido.

De ser así, puede que la persona en concreto ya no pueda ver sus legítimas expectativas colmadas e individualmente obtener una verdadera tutela efectiva. Sin embargo, la colectividad a la que pertenece, las personas en las que concurre esa misma circunstancia por la que se le ha discriminado, como colectivo, sí pueden ver amparados sus derechos.

Esa nueva tutela que, sin renunciar a la sanción, va un paso más allá y se sitúa también en la prevención de la discriminación de una colectividad, es la base del Plan de Igualdad y de la recientemente adoptada Ley Orgánica 2/2024, de Paridad. Por centrarme en lo primero, lo que caracteriza al Plan de Igualdad como instrumento garante de la igualdad de mujeres y hombres es iniciar un diagnóstico con el que se analiza la realidad de la empresa desde la perspectiva de la discriminación indirecta y directa y, de entender que concurre discriminación, y en función de en qué medida se da esa discriminación, adoptar unas medidas u otras para erradicar esa discriminación desde una obligación de hacer en un determinado sentido.

Es decir, si una empresa recurre a un algoritmo y éste nunca selecciona, entre posibles candidatos de ambos sexos y con idénticos conocimientos, a una mujer, siguiendo la empresa en su contratación la sugerencia del algoritmo, esta realidad se podrá evidenciar, de existir, en el Plan de Igualdad. Bien de forma previa a su implantación, en el diagnóstico previo; bien, implantado este, en las reuniones de la comisión de control y

seguimiento del mismo. Se evidenciará la posible discriminación, se analizará la posible existencia o no de una causa objetiva que pueda justificar la ausencia de mujeres entre los candidatos seleccionados y, de no ser así, confirmada la discriminación, se adoptarán medidas para erradicarla.

Del mismo modo, y desde el mismo razonamiento lógico, si aplicando ese mismo algoritmo en la selección de consejeros o personal de alta dirección de las empresas cotizadas o las entidades de interés público en los términos que exige la Ley Orgánica 2/2024, de Paridad, no se cumplen con los porcentajes de presencia equilibrada que marca la ley, habrá que rehacer la selección hasta conseguir la presencia equilibrada en los términos que marca el hoy art. 9 de la mencionada Ley Orgánica 2/2024, de Paridad.

En los ejemplos expuestos, que tienen como común denominador que sea el sexo de los candidatos el motivo por el cual se discrimina, la evolución normativa y los instrumentos jurídicos diseñados en favor de la igualdad real de mujeres y hombres, podrán ayudar a evidenciar que un proceso de selección algorítmica, en un caso determinado, incurre en discriminación. Lo que desde la perspectiva indirecta ocurrirá siempre que genere sin motivo objetivo justificado un efecto adverso en un colectivo respecto otro. Siendo indiferente, a efectos discriminatorios, que el efecto adverso lo produzca la fórmula matemática que ha sido utilizada en el diseño del algoritmo, los datos de los que se nutre el algoritmo o *esa caja negra* a la que difícilmente se tiene acceso[19].

Pero claro, como se ha expuesto en los dos epígrafes anteriores, ni toda la discriminación es solo directa o indirecta; ni tampoco toda discriminación se produce por razón de sexo, aunque muchas sí. Y, en estas otras circunstancias y en estas otras conductas discriminatorias, lo determinante será adquirir conciencia, primero, y demostrar, después, que la selección algorítmica ha incurrido en discriminación.

Pensemos por ejemplo en el concepto de la discriminación por asociación o por error. Fundamentalmente en esta última manifestación de discriminación que, como se exponía, desde el art. 6.2 de la Ley 15/2022 es definida como *aquella que se funda en una apreciación incorrecta acerca de las características de la persona o personas discriminadas.*

Pues bien, una selección mediante algoritmos o sistemas de IA, sin duda, puede generar muchas discriminaciones por error. Piénsese que, en muchas ocasiones, lo que se hace en estos procesos selectivos es atribuir características a unas personas tras realizar un estudio estadístico de probabilidades en función del comportamiento de otras personas distintas. Presumiendo, tras ese análisis estadístico y una cierta correlación de premisas, que, si concurren determinadas cualidades o características en una persona, ésta va a comportarse de una determinada manera. Lo que al final supone que se la elige o no para

[19] En sentido contrario, haciéndose eco de parte de la doctrina norteamericana que cuestiona la aplicabilidad del *disparate impact* a los procedimientos donde intervienen los algoritmos por poner el acento en el parámetro o la causa del efecto adverso, véase, SAEZ LARA, C., "El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva…", cit., págs. 48-49. No obstante, como digo, no considero que esa doctrina sea aplicable a estos efectos, porque en nuestro ordenamiento jurídico también las prácticas que generan un efecto adverso, aunque se desconozcan las concretas causas o parámetros, se incluyen en el concepto legal de discriminación indirecta.

un puesto de trabajo según el algoritmo realice ciertas deducciones en base al comportamiento de otras personas.

Lo anterior es lugar común en el mundo de la IA. Es de sobra conocido que, uno de los recursos del análisis en selección de personal es la filmación de entrevistas de trabajo. En estos casos, se presume que, tras informar de la grabación a la que están siendo sometidos los candidatos, la IA se emplea para evaluar tanto las señales conscientes e inconscientes de las personas candidatas en su entrevista de trabajo[20]. Y ello pese a que, en nuestro país, la "Carta de Derechos Digitales" limita, aunque ciertamente no prohíbe, los "sistemas de análisis de personalidad o conducta que impliquen la toma de decisiones automatizadas o el perfilado de individuos, o grupos de individuos" (art. V)[21].

En todo caso, la práctica anterior es algo bastante común. Y aunque las deducciones y presunciones que pueda extraer el algoritmo del estudio del comportamiento humano, basada en el modelo estadístico, en muchas ocasiones serán claramente eficaces, no siempre acertaran en sus presunciones. Cuando erren, pueden ocasionar no solo un perjuicio a la persona afectada, y una vulneración de ciertos derechos fundamentales, sino incurrir en una discriminación por error, lo que evidentemente ocurrirá cuando tras esas deducciones o premisas se excluye de la selección a una persona por otorgarle de forma equivocada, tras un análisis estadístico, unas cualidades en base a ciertos comportamientos propios o ajenos que acaban siendo la causa de la discriminación.

Y claro, lo peor, es que el error en esos casos es también una prueba diabólica. La discriminación por error, en realidad, y en cuanto tal, en un proceso selectivo, aun pudiéndose producir, será difícilmente demostrable y, por tanto, controlable[22].

La discriminación en el proceso de selección ha existido desde siempre. De hecho, simplificando mucho el problema, y como se apuntaba, la discriminación genera perjuicios por prejuicios. Y las personas humanas, también las encargadas de los procesos de selección, tenemos ciertos prejuicios. Por lo que el riesgo de que se produzca una discriminación en el proceso de selección existe, con y sin algoritmo. Y, desde luego, la preocupación al respecto es previa a la irrupción masiva de los algoritmos y la IA en los procesos selectivos[23]. Además, recurrir al algoritmo y a la IA en el momento de la selec-

[20] Uno de esos productos, fabricado por HireVue, está funcionando en más de seiscientas compañías. Cfr.https://www.bbvaopenmind.com/articulos/inteligencia-artificial-en-entorno-laboral-desafios-para-trabajadores/ 16 de agosto de 2024.

[21] https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf. 17 de agosto de 2024.

[22] Cuestión distinta es que la discriminación por error se pudiese detectar gracias a una discriminación indirecta. Por dejar de ser individual y pasar a ser grupal. Pero cuando sea discriminación por error individual, será muy difícil de demostrar.

[23] Recuérdese el desgraciado y famoso incidente que afectó a una compañía de supermercados y en el que se denunciaba que habían aparecido currículos con anotaciones discriminatorias en una bolsa de basura. La compañía negó rotundamente los hechos, y la Inspección de Trabajo de Madrid inició una investigación y la Agencia de Protección de Datos concluyó que la empresa había incurrido en una infracción grave al no proteger datos personales. También es llamativo el caso de la SSTSJ

ción no está prohibido, es lícito y legal. Lo que ocurre es que la preocupación aumenta cuando se adquiere conciencia del riesgo que existe en los supuestos en los que el proceso de selección se instrumentaliza, por cuanto que el algoritmo y la IA pueden aumentar y expandir exponencialmente la discriminación preexistente en la sociedad, incluso a veces sin intencionalidad.

La opacidad que caracteriza a la IA y a los algoritmos incrementa la complejidad de fiscalizar, controlar, si en un procedimiento selectivo en concreto en el que se selecciona a una o unas personas, y no a otra u otras, se ha incurrido o no en discriminación.

Empezando por la difusión de las ofertas de empleo, que puede realizarlas un algoritmo casi de manera automática incurriendo ya, en ese primer momento, en una conducta discriminatoria; pasando por el propio proceso selectivo en lo que a pruebas o exámenes a las personas candidatas se refiere[24]. Sin ignorar que en estos procesos se pueden valorar conductas conscientes o inconscientes de las personas candidatas de las que se pueden extraer, con razón o sin ella, determinadas conclusiones respecto las que adoptar la discriminatoria decisión de no elegir a alguien por las deducciones algorítmicas[25].

Lo anterior partiendo de la premisa de que la empresa que utiliza el algoritmo o recurre a una tercera para un proceso de selección que a su vez utiliza el algoritmo, conoce los parámetros, premisas o reglas de actuación del algoritmo. Pero puede que no sea así. Los algoritmos o sistemas de inteligencia artificial son cada vez más complejos y no atienden exclusivamente a elementos predefinidos, sino que aprenden por sí mismos[26]. La persona encargada de la selección, por tanto, puede solicitarle al algoritmo que seleccione a los candidatos en función de su eficacia o idoneidad, pero, no obstante, desconocer las variables que éste ha tenido en cuenta para optar por una persona y no por otra, imposibilitándose así el control de una posible discriminación de forma previa al análisis de sus resultados.

---

Cataluña de 8 de marzo de 1999, Ar.2552, que aludiendo a la neutralidad de la tercera persona del singular no considera discriminatoria una oferta de empleo que decía lo siguiente "se necesita chico dinámico, libre del servicio militar (…)".

[24] Como se viene apuntando, "hay plataformas muy sofisticadas que recolectan y analizan los datos de quienes se postulan a puestos de trabajo en una o más empresas, y posteriormente automatizan el proceso de seleccionar currículums a partir de criterios que pueden estar determinados. Estas plataformas o PARS suelen ser utilizados por empresas especializadas en el rubro y se enfocan en la elaboración de una lista corta de potenciales candidatos y en la realización de exámenes, juegos o pruebas, dejando, en ocasiones, las etapas finales a cargo de la empresa contratante (CODDOU MCMANUS, A. y PADILLA PARGAR, R., "Discriminación algorítmica en los procesos automatizados de reclutamiento y selección de personal". Revista chilena de derecho y tecnología, n.º 13, 2024, págs. 186-219)".

[25] Por ejemplo, un comportamiento amanerado o afeminado en un hombre no necesariamente comporta una determinada orientación o identidad sexual. Igual que determinada apariencia no siempre determina una situación socioeconómica, etc. Circunstancias, las expuestas, consideradas como causa de discriminación expresamente en el art. 2.1. de la Ley 15/2022.

[26] Cfr. TODOLÍ SIGNES, A. Algoritmos productivos y extractivos. Aranzadi. Pamplona, 2023.

## 5. Los medios de tutela ante el riesgo de discriminación en la selección algorítmica

Evidentemente la norma no es ajena a todas las posibles discriminaciones expuestas. Y no lo es ni en la norma generalista ni tampoco en la específica referente a la IA.

Si bien hasta ahora me he centrado básicamente en identificar algunos de los riesgos que comporta esta revolución tecnológica y digital desde el punto de vista de la discriminación en la selección por algoritmos o IA, centrándome en los conceptos de discriminación jurisprudencial y normativa, señalado como posibles instrumentos específicos de tutela al respecto los existentes en materia antidiscriminatoria, sobre todo por razón de sexo, como los Planes de Igualdad o las concretas exigencias de presencia equilibrada que exige la Ley Orgánica 2/2024, de Paridad, que, al menos en el caso de la discriminación por sexo, servirán para detectar un posible incumplimiento empresarial en un proceso de selección pese a no conocer la trazabilidad del algoritmo, lo cierto es que también desde el punto de vista estricto de la IA se han intentado establecer en la normativa genérica (ET) y específica (antidiscriminatoria general y de protección de datos e IA) una serie de garantías para mitigar los posibles riesgos discriminatorios que entraña recurrir a un algoritmo o a la IA en los procesos de selección.

Desde un punto de vista más generalista, así, el ET se modificó en su art. 64.4.d[27] para reconocer el derecho del comité de empresa a ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles. Información que, parece, para cumplir con el mandato legal deberá ser lo suficientemente explícita y accesible. No pareciendo que se cumpla con el mandato legal con la mera puesta a disposición a la RLT de una información ininteligible o incomprensible[28].

Mucho se ha escrito al respecto de este deber de informar[29]. Se discute si es un derecho estático o dinámico. Y, sobre todo, la efectividad práctica que puede tener este derecho a informar como garante de la tutela de los trabajadores individualmente

[27] La redacción actual es la del art. único.1 de la Ley 12/2021, de 28 de septiembre, por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales, que ratifica la que le dio el Real Decreto-ley 9/2021, de 11 de mayo, por el que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales.

[28] Cfr. GINÈS I FABRELLAS, A., "El tiempo de trabajo en plataformas: ausencia de jornada mínima, gamificación e inseguridad algorítmica". LABOS Revista de Derecho del Trabajo y Protección Social, n.º 1, 2021, p. 38.

[29] Por todos, CRUZ VILLALÓN, J.: "La participación de los representantes de los trabajadores en el uso de los algoritmos y sistemas de inteligencia artificial" http://jesuscruzvillalon.blogspot.com/2021/05/; GÓMEZ GORDILLO, R., "Algoritmos y derecho de información de la representación de las personas trabajadoras". Temas laborales, n.º 158, 2021, págs. 178 y ss.

afectados por la aplicación de un algoritmo[30]. El secreto industrial por parte de la empresa y la en ocasiones difícil trazabilidad de los algoritmos son algunos de los motivos a los que se alude para advertir que la eficacia de esta medida seguramente será, en el sentido expuesto, menor de la originariamente esperada. Por el contrario, parece claro que es una obligación empresarial que no acepta excusas y que la información que se debe trasladar debe permitir conocer a la RLT donde se utiliza el algoritmo y cómo funciona. Seguramente, más que dar a conocer la concreta fórmula matemática que en su caso pudiese estar protegida por el secreto industrial, habrá que informar sobre que variables se toman en consideración en cada caso por los algoritmos para tomar una decisión y no otra.

Con todo, el art. 64.4.d. ET habla de informar al respecto de *los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial,* cuando no hay que olvidar que a los efectos que aquí interesan tan importantes como estos son los datos con los que los algoritmos o el sistema de IA han sido entrenados, o de los que se nutren. Y estos dudosamente quedan incluidos en el mencionado precepto legal[31].

Por lo demás, desde el punto de vista de la normativa más específica de entre las normas antidiscriminatorias, seguramente por ser la más reciente en el tiempo, es la Ley 15/2022, la que expresamente se refiere a la IA en su art. 23, aunque poco determinante para la empresa privada. La previsión legal para con la empresa privada es mucho menos expansiva que para con las Administraciones Públicas. De hecho, sorprende, cuanto menos, que si bien a las empresas privadas solo se les menciona para exigirles que *promuevan el uso de una Inteligencia Artificial ética, confiable y respetuosa con los derechos fundamentales, siguiendo especialmente las recomendaciones de la Unión Europea en este sentido,* a las AAPP se les imponga, además de lo expuesto, otros mandatos más expeditivos como el de transparencia en el diseño y en la implementación así como capacidad de interpretación de las decisiones adoptadas por los algoritmos además de *favorecer la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio.*

Como mencionaba, resulta llamativo que en una norma tan transversal y con una regulación tan extensa y ambiciosa como la Ley 15/2022, se contengan previsiones para con las AAPP y no para con la empresa privada. Dejando pasar la oportunidad de fijar verdaderos instrumentos de prevención y tutela frente al riesgo que la IA puede generar desde la perspectiva de la discriminación, fundamentalmente, aunque no exclusivamente desde la perspectiva laboral, en el acceso al empleo.

---

[30] RODRIGUEZ CARDO, I. "Decisiones automatizadas y discriminación algorítmica…", cit., págs. 18 y ss.

[31] Recuérdese el caso de discriminación por razón de sexo en Amazon, antes mencionado.

En todo caso, como ya se ha expuesto, lo anterior en absoluto debe interpretarse como que la empresa privada queda al margen de la responsabilidad de garantizar los derechos fundamentales y la prohibición de discriminación cuando seleccionan a los candidatos mediante algoritmos o fórmulas de IA. Aunque no lo diga expresamente la Ley 15/2022, la transparencia o la información de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial ya resulta exigible, al menos respecto la RLT en el art. 64.4.d ET; y la evaluación del impacto que la Ley 15/2022 exige para con las AAPP puede deducirse como obligación para con la empresa privada tanto del Reglamento General de Protección de Datos como del recientemente aprobado y publicado Reglamento de Inteligencia Artificial.

En efecto, así es. De hecho, el primero de ellos, que también fue el primero de los dos desde el mero punto de vista cronológico, en sus arts. 22, 32 y 35 exige que en todo tratamiento automatizado realizado en base a la IA se garantice el "humano al mando"[32] además de comportar una evaluación de impacto.

El segundo, más reciente, y un poco más explícito, si bien también menos protector para con la tutela de derechos de los destinatarios de lo que se esperaba, califica la selección de personal y gestión de las relaciones laborales como sistema de alto riesgo, reiterando la exigencia, por tanto, de control humano (humano al mando), del deber de información y de transparencia; y, al responsable del despliegue, de la evaluación de impacto.

Y esta última sí puede ser determinante para verificar si se produce o no la discriminación con un determinado sistema de IA o algoritmo en un proceso de selección, al permitir analizar los resultados.

---

[32] Aunque tras la aprobación del RIA y cuanto allí se exige la discusión puede entenderse superada por la consideración de la gestión de personal como sistema de alto riesgo, es importante señalar que una parte de la doctrina ha concluido que el art. 22 RGPD y el requisito del humano al mando no tiene una aplicación tan automática al ámbito jurídico laboral como se ha dicho. En ese sentido, RODRIGUEZ CARDO, I. "Decisiones automatizadas y discriminación algorítmica…", cit., defiende "que el art. 22 RGPD pretende en realidad proteger al consumidor o usuario, y no tanto al trabajador, y, lógicamente, no cabe extender sin la debida modulación el marco normativo diseñado para la relación entre un productor o prestador de servicios con un potencial cliente (o usuario o consumidor) a una relación tan distinta como es la que une a empresario y trabajador. El RGPD pretende evitar, por ejemplo, que una compañía de seguros descarte a priori la contratación con determinados perfiles de personas por el hecho de que estadísticamente suponen un riesgo (menor esperanza de vida). En cambio, en el contexto laboral no puede hurtarse al empleador la capacidad de determinar qué habilidades y competencias profesionales resultan más valiosas para desarrollar determinadas funciones, por lo que las decisiones automatizadas y la elaboración de perfiles profesionales son, en principio, opciones lícitas, sin perjuicio de que no resulte admisible el tratamiento de ciertos datos que el empleador no está legitimado para conocer o procesar". En el mismo sentido, GARRIGA DOMÍNGUEZ, A., "La elaboración de perfiles y su impacto en los derechos fundamentales. Una primera aproximación a su regulación en el Reglamento General de Protección de Datos de la Unión Europea", Derechos y Libertades: revista de filosofía del derecho y derechos humanos, n.º 38, 2018, págs. 107 y ss. En sentido contrario, PRECIADO DOMENECH, C.H., "Algoritmos y discriminación en la relación laboral", Jurisdicción Social, n.º 223, 2021.

Es verdad, con todo, que la aprobación del RIA es demasiado reciente como para precisar la efectividad que estas garantías van a tener en el destinatario final en cuanto a la protección de la no discriminación. Por lo pronto, se evidencia la preocupación del RIA por poner a disposición del usuario, en este caso, del empleador, el conocimiento, entre otras cosas, de cualquier circunstancia conocida o previsible asociada al uso del sistema de IA conforme a su finalidad prevista o a un uso indebido razonablemente previsible, que pueda dar lugar a riesgos para la salud y la seguridad o los derechos fundamentales; al tiempo que, al candidato afectado, se le informará de que está siendo examinado por un algoritmo o un sistema de IA[33].

[33] Art. 13 RIA. 1. Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de forma que se garantice que su funcionamiento es lo suficientemente transparente como para permitir a los implantadores interpretar los resultados del sistema y utilizarlos adecuadamente. Se garantizará un tipo y un grado de transparencia adecuados con vistas a lograr el cumplimiento de las obligaciones pertinentes del proveedor y del implantador establecidas en la sección 3.

2. Los sistemas de IA de alto riesgo irán acompañados de instrucciones de uso en un formato digital adecuado o de otro tipo que incluya información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible para quienes los desplieguen.

3. Las instrucciones de uso contendrán como mínimo la siguiente información:

(a) la identidad y los datos de contacto del prestador y, en su caso, de su representante autorizado;

(b) las características, capacidades y limitaciones de rendimiento del sistema de IA de alto riesgo, incluyendo:

(i) su finalidad prevista;

(ii) el nivel de precisión, incluidas sus métricas, solidez y ciberseguridad a que se refiere el artículo 15, con respecto al cual se ha probado y validado el sistema de IA de alto riesgo y que cabe esperar, así como cualquier circunstancia conocida y previsible que pueda repercutir en ese nivel previsto de precisión, solidez y ciberseguridad;

(iii) cualquier circunstancia conocida o previsible, relacionada con el uso del sistema de IA de alto riesgo de conformidad con su finalidad prevista o en condiciones de uso indebido razonablemente previsible, que pueda dar lugar a riesgos para la salud y la seguridad o los derechos fundamentales a que se refiere el artículo 9, apartado 2;

(iv) en su caso, las capacidades técnicas y las características del sistema de IA de alto riesgo para proporcionar información pertinente para explicar sus resultados;

(v) cuando proceda, su rendimiento en relación con las personas o grupos de personas específicos sobre los que se pretende utilizar el sistema;

(vi) cuando proceda, especificaciones de los datos de entrada, o cualquier otra información pertinente en cuanto a los conjuntos de datos de entrenamiento, validación y prueba utilizados, teniendo en cuenta la finalidad prevista del sistema de IA de alto riesgo;

(vii) en su caso, información que permita a los responsables del despliegue interpretar el resultado del sistema de IA de alto riesgo y utilizarlo adecuadamente;

(c) los cambios en el sistema de IA de alto riesgo y en su funcionamiento que hayan sido predeterminados por el proveedor en el momento de la evaluación inicial de la conformidad, en su caso;

(d) las medidas de supervisión humana contempladas en el artículo, incluidas las medidas técnicas establecidas para facilitar la interpretación de los resultados de los sistemas de IA de alto riesgo por parte de quienes los despliegan;

Con todo, es difícil apreciar la obligación de informar al candidato por parte de la empresa de los parámetros que ha utilizado el algoritmo para seleccionarle o para no hacerlo. Será informada la RLT; y será informado el empleador como usuario en los términos referidos. Pero no está tan claro, en base al RIA, que la persona cuyo curriculum o comportamiento sea analizada por IA en aras de ser contratado o no, deba ser informado por la empresa de forma precisa y minuciosa de los distintos aspectos que se han baremado para considerarlo o no apto para un puesto de trabajo.

Y lo mismo respecto la evaluación de impacto. Está claro que en base al RIA debe hacerse por el responsable del despliegue una evaluación de impacto, y que esta es positiva para apreciar la posible incurrencia en discriminación, aunque sea a posteriori. Pero la evaluación de impacto, tal y como está prevista, no creo que se traduzca en la obligación de informar fehacientemente, al menos a priori, a personas que son candidatas, pero no trabajadoras, del impacto que ha tenido el algoritmo entre mujeres y hombres, mayores y jóvenes, por estado de salud, por raza, opinión, ideología, origen socioeconómico, etc.

Lo que al final, básicamente, requiere volver a insistir en la relevancia de interpretar integradamente la normativa específica de IA con la antidiscriminatoria. Que, si bien en materia de discriminación por razón de sexo no generará tantos problemas, si la empresa implanta un plan de igualdad o se ve afectada en consejeros y alta dirección por la LO 2/2024, de Paridad, exigirá estar muy pendientes en otro tipo de discriminación como la que se puede dar en caso de que el candidato presente alguna discapacidad, alteración funcional o sea tratado discriminatoriamente por cualquier causa constitucional y legalmente prohibida.

En todo caso, veremos como va avanzando la aplicación de la RIA. Es demasiado pronto, para extraer conclusiones definitivas en un tema tan complejo como fragmentado normativamente.

## 6. A modo de recapitulación

Llegados a este punto, y siendo el momento de concluir, las ideas que deben destacarse como recapitulación de lo expuesto pueden sistematizarse como sigue:

*Primero.* La selección mediante algoritmos o IA en nuestro ordenamiento jurídico es una opción empresarial perfectamente lícita. El empleador es quien decide si realiza el proceso de selección mediante métodos tradicionales o recurriendo a algoritmos o IA.

(e) los recursos informáticos y de hardware necesarios, la vida útil prevista del sistema de IA de alto riesgo y las medidas de mantenimiento y cuidado necesarias, incluida su frecuencia, para garantizar el correcto funcionamiento de dicho sistema de IA, incluso en lo que respecta a las actualizaciones de software;

(f) cuando proceda, una descripción de los mecanismos incluidos en el sistema de IA de alto riesgo que permita a los responsables de la aplicación recopilar, almacenar e interpretar adecuadamente los registros de conformidad con el artículo 12.

En todo caso, la opción es empresarial. Por lo que tanto la decisión de contratar a una persona en particular como la de no hacerlo también lo son. La empresa, en caso de incurrir en prácticas ilícitas, no puede escudarse en el recurso al algoritmo, que carece de personalidad jurídica y, en consecuencia, no es titular ni de derechos ni de obligaciones.

*Segundo.* El proceso de selección y la decisión de contratar a una persona u a otra, mediando o no algoritmo o IA, forma parte de las potestades empresariales que emanan del art. 38 CE. Que, en cuanto tal, y dado que se trata de una potestad discrecional que no arbitraria, deberá someterse al obligado respeto a los derechos fundamentales de las personas candidatas y a la prohibición de incurrir en discriminación.

*Tercero.* Asentada en los tribunales la interpretación integrada del art. 14 CE y del art. 9.2. CE, en favor del principio de igualdad real y la prohibición de discriminación, la evolución de este último concepto, en la jurisprudencia, primero, y en la ley, después, exige considerar distintas manifestaciones de la prohibición de discriminación a efectos de valorar fehacientemente cuando un algoritmo o un sistema de IA cumple con los parámetros constitucionales y legales. O, cuando, pretendiéndolo o no, estos límites al poder de selección empresarial han sido vulnerados.

*Cuarto.* Seguramente, los casos de discriminación indirecta serán, por las características del proceso de selección, los más fáciles de detectar y demostrar. Pero no por ello deben dejar de considerarse, en estos procedimientos, la posible incurrencia en otras prácticas discriminatorias. Desde las más evidentes como la discriminación directa o las más fácilmente demostrables como la indirecta, a las más complejas como la discriminación por error, por asociación, múltiple o interseccional.

*Quinto.* Claro, las características del algoritmo o de la IA –volumen de bancos de datos, secretismo de la fórmula matemática, caja negra, algoritmos no solo deductivos sino inductivos de difícil trazabilidad– incrementan exponencialmente los riesgos de que se produzcan sesgos y decisiones discriminatorias incluso a veces de manera involuntaria. En la realidad existe y persiste la discriminación. Lo que hay que evitar, por tanto, es que el algoritmo o la IA, en lugar de contribuir a erradicarla, la magnifique.

*Sexto.* Indudablemente la norma no es ajena a lo expuesto. Lo que ocurre es que las garantías previstas para evitar la discriminación en el proceso de selección instrumentalizado mediante un algoritmo o IA no parece que siempre vayan a ser lo eficaces que cabría esperar. Dejando al margen la discriminación por razón de sexo, en el que los propios instrumentos previstos en las específicas normas antidiscriminatorias, sí pueden revelarse como un complemento efectivo en la tutela antidiscriminatoria desde el análisis de los resultados, en el resto de supuestos habrá que confiar fundamentalmente en las previsiones del ET y la normativa específica en IA, esto es, en que al respecto resulten efectivos el deber de información a la RLT en los términos del art. 64.4.d ET, y el control humano o

humano al mando, la transparencia e información y la evaluación de impacto. Todo ello en los términos del RGPD y del muy reciente RIA.

*Séptimo.* La regulación en esta materia es todavía incipiente. Tanto que, al menos por lo que afecta a la RIA, entiendo contraproducente realizar aseveraciones contundentes y firmes en cuanto a su relativa eficacia en lo que aquí interesa. Seguramente si la AESIA, que tiene apenas un mes de antigüedad, cumple con el papel asignado, la confluencia de todas las medidas empiece a ser relevante en favor de la tutela antidiscriminatoria.

*Octavo.* En cualquier caso, el objetivo de este trabajo no es en absoluto demonizar la IA ni los algoritmos, sino todo lo contrario. Lo que se pretende con estas reflexiones es contribuir a su verdadera eficacia. Y la mejor forma de hacerlo, en un tema tan complejo como el de la selección de personal para el acceso a la empresa, es señalando el riesgo de incurrir o reproducir ciertos prejuicios preexistentes que, al margen de discriminar a una persona o grupo de personas con un denominador común, ignoren apriorísticamente la potencialidad de una parte de la sociedad cuando pretende acceder a un empleo.

## 7. Bibliografía

BELTRÁN DE HEREDIA RUIZ, I., "Algoritmos psicometría y derechos del yo inconsciente de la persona en el ámbito socio-laboral". Revista Derecho Social y Empresa, nº.18, 2023.

CRUZ VILLALÓN, J., "La participación de los representantes de los trabajadores en el uso de los algoritmos y sistemas de inteligencia artificial" http://jesuscruzvillalon.blogspot.com/2021/05/.

GARRIGA DOMÍNGUEZ, A., "La elaboración de perfiles y su impacto en los derechos fundamentales. Una primera aproximación a su regulación en el Reglamento General de Protección de Datos de la Unión Europea". Derechos y Libertades: revista de filosofía del derecho y derechos humanos, n.º 38, 2018, págs. 107 y ss.

GINÈS I FABRELLAS, A., "El tiempo de trabajo en plataformas: ausencia de jornada mínima, gamificación e inseguridad algorítmica". LABOS Revista de Derecho del Trabajo y Protección Social, n.º 1, 2021, págs. 19-42.

GINÈS I FABRELLAS, A., "Sesgos discriminatorios en la automatización de decisiones en el ámbito laboral: evidencias de la práctica". En RIVAS VALLEJO, P., *Discriminación algorítmica en el ámbito laboral: perspectiva de género e intervención*. Aranzadi. Pamplona, 2022.

GINÈS I FABRELLAS, A., *Algoritmos, inteligencia artificial y relación laboral*. Aranzadi. Pamplona, 2023.

GÓMEZ GORDILLO, R., "Algoritmos y derecho de información de la representación de las personas trabajadoras", Temas laborales, n.º 158, 2021, págs. 178 y ss.

MERCADER UGUINA, J. R., "Discriminación algorítmica y derecho granular: nuevos retos para la igualdad en la era del Big data". LABOS Revista De Derecho Del Trabajo y Protección Social, nº2, 2021, págs.4-10.

OLARTE ENCABO, S., "La aplicación de inteligencia artificial a los procesos de selección de personal y ofertas de empleo. Impacto sobre el derecho a la no discriminación". Documentación Laboral, n.º 119, 2020, págs. 95-97.

PEYRONNET, M., "El uso de los algoritmos y la inteligencia artificial en la selección de personal: ¿una ocasión para eliminar la discriminación". En MOLINA NAVARRETE, C., y VALLECILLO GÁMEZ, M. R. (Dir.), *De la economía digital a la sociedad del E-Work decente: condiciones sociolaborales para una industria 4.0 justa e inclusiva*. Aranzadi, Pamplona, 2021.

PRECIADO DOMENECH, C.H., "Algoritmos y discriminación en la relación laboral". Jurisdicción Social, n.º 223, 2021.

RODRIGUEZ CARDO, I., "Decisiones automatizadas y discriminación algorítmica en la relación laboral ¿hacia un Derecho del Trabajo de dos velocidades?". Revista Española de Derecho del Trabajo, n.º 253, 2022, págs. 135-188.

SAEZ LARA, C., "El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación". Temas laborales: Revista andaluza de trabajo y bienestar social, n.º 155, 2020, págs. 42-60.

TODOLÍ SIGNES, A. Algoritmos productivos y extractivos. Aranzadi. Pamplona, 2023.

VIOLLIER, P., y VELASCO, P., "El uso de toma de decisiones automatizadas para la selección de personal". En SEVERÍN CONCHA, J.P. (Ed.), *Derechos fundamentales de la persona del trabajador*. Tirant lo Blanch, Valencia, 2021.

# Sistemas de inteligencia artificial y prevención de los riesgos laborales. Obligaciones del proveedor y del empresario

## Artificial intelligence systems and the prevention of occupational risks: obligations of the supplier and the employer

José Luis Goñi Sein
*Catedrático de Derecho del Trabajo y Sseguridad Social*
*Universidad Pública de Navarra*
ORCID ID: 0000-0003-4481-9483

doi: 10.20318/labos.2024.9036

*Resumen:* El propósito de este trabajo es analizar la interación existente entre el Reglamento UE de Inteligencia Artificial y la normativa de prevención de riesgos laborales, respecto de la utilización de los sistemas de IA en el ámbito laboral. Estas dos normativas presentan una identidad de razón en cuanto al enfoque basado en el riesgo y al objetivo, pues tratan de mitigar los potenciales riesgos, pero, al mismo tiempo, presentan diferencias notables sobre el alcance del riesgo, dando lugar a antinomias. Se trata de resolver estas antinomias, haciendo una interpretación integradora de los dos sistemas normativos y de sus finalidades, con el objeto de establecer el conjunto de obligaciones que deben tener en cuenta, tanto el proveedor de IA, como el responsable del despliegue (empresario) para su introducción en el mercado y su puesta en servicio, desde el punto de vista de la prevención de riesgos laborales.

*Palabras clave:* Inteligencia artificial, ámbito laboral, riesgos laborales, proveedor de IA, empresario, obligaciones preventivas.

*Abstract:* The purpose of this paper is to analyse the interplay between the EU Regulation on Artificial Intelligence and the EU regulation on occupational risk prevention with regard to the use of AI systems in the workplace. These two regulations present an identity of reason in terms of risk-based approach and objective, as they seek to mitigate potential risks, but, at the same time, they present notable differences on the scope of risk, giving rise to antinomies. The aim is to resolve these antinomies by making an integrative interpretation of the two regulatory systems and their purposes, in order to foresee which set of obligations must be taken into account, both by the AI provider and the person responsible for the deployment (employer) for its introduction in the market and its commissioning, from the point of view of occupational risk prevention.

*Keywords:* Artificial intelligence, work environment, occupational risks, AI provider, employer, preventive obligations.

## 1. Introducción

La IA aporta numerosos beneficios a la vida de los ciudadanos, también al ámbito de la seguridad y salud laboral, ayudando a identificar los peligros, "predecir riesgos potenciales, proporcionar monitoreo en tiempo real, reconocer comportamientos inseguros, detectar condiciones inseguras y sugerir formas de riesgos potenciales"[1]. Asimismo, resulta muy útil para conocer el estado de salud o mejorar la vigilancia de la salud del trabajador, evaluar el grado de malestar psicológico, o los posibles efectos negativos para la salud mental del trabajador, en particular, cuando está sometido a presión para alcanzar un determinado nivel de productividad. Mediante sistemas de gestión de personal basada en la IA cabe detectar el grado de estrés, el síndrome de desgaste profesional o y de agotamiento, de manera que se puedan adoptar medidas de prevención[2].

Pero no son las capacidades o beneficios de la IA el objeto de este análisis, sino los desafíos que para las personas y la sociedad presenta la incorporación de la IA al ámbito de las relaciones laborales, en concreto, a los equipos de trabajo y sistemas de gestión de recursos humanos. El desafío principal, dejando aparte los problemas jurídicos, o éticos, es que los sistemas de IA o las máquinas u ordenadores que incorporen dichos sistemas, no funcionen correctamente o sencillamente generen riesgos, convirtiéndose en daños en el mundo laboral para la seguridad y salud laboral, incluyendo los riesgos psicosociales. Y ello no solo como consecuencia de errores en el diseño y fabricación de los sistemas de IA, porque pueden contener "errores significativos y alucinaciones debidas a circunstancias multifactoriales como la desactualización de los datos de entrenamiento, fallos informáticos, intereses comerciales, sesgos por género, raza y contextos sociales"[3], sino porque sus actos resultan menos predecibles en virtud de que pueden actuar y perseguir objetivos de forma autónoma.

La identificación y el control de esos riesgos para los derechos, la seguridad y el buen funcionamiento del mercado único de la Unión Europea constituyen el punto central y una de las aristas más complejas de la IA. La determinación del riesgo es el *prius* que condiciona los parámetros, bien de uso inaceptable o de uso legítimo del sistema

[1] Vid. EU-OSHA: El impacto de la Inteligencia Artificial en la Seguridad y Salud en el trabajo". Disponible en: https://osha.europa.eu/es/publications/impact-artificial-intelligence-occupational-safety-and-health.

Gracias a la creciente disponibilidad de datos y macrodatos (big data) y a la capacidad de usar algoritmos para el tratamiento de datos, los sistemas de IA permiten monitorear continuamente los parámetros de las máquinas y los entornos de trabajo, lo que ayuda a reducir el riesgo de errores mecánicos y humanos. Los sistemas de IA "pueden entrenarse para detectar peligros de seguridad que los humanos tal vez no puedan ver, como grietas microscópicas en una estructura o cambios sutiles en los vitales de un paciente" vid. ABDULLAH MALIK: *Artificial Intelligence in Health and Safety*, 22 de febrero de 2023 Disponible en: https://safetypedia-com.translate.goog/safety/artificial-intelligence-in-health-and-safety/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc.

[2] Sobre dispositivos inteligentes para la seguridad y salud en el trabajo, vid. LLORENS ESPADA, J.*: Límites al uso de la Inteligencia Artificial en el ámbito de la salud*, La Ley, Madrid, 2023, pp. 151 y ss.

[3] VESTRI, G.: "La Unión Europea estrena el Reglamento de Inteligencia Artificial (RIA).Control, supervisión y uso de una tecnología cada vez más presente en la vida de todos", *Diario LA LEY*, Nº 10550, 19 de Julio de 2024.

de IA, porque se quiere que la IA sea desarrollada y utilizada de manera segura, ética y respetuosa con los Derechos fundamentales y los valores fundacionales de la Unión Europea Las obligaciones impuestas a los proveedores o usuarios vendrán determinadas en función de los riesgos concretos que comporta el uso de la IA.

Pero la noción de riesgo que se toma en consideración en este estudio es un riesgo algo más concreto, está relacionada con la seguridad y la salud de la persona del trabajador en el ámbito de la relación laboral. La perspectiva que aquí interesa es la de los riesgos conocidos y razonablemente previsibles de los sistemas de IA para la salud, la seguridad de los trabajadores (dejando fuera los relativos a los derechos fundamentales), teniendo en cuenta su finalidad prevista y también su uso indebido razonablemente previsible, así como los posibles riesgos derivados de la interacción entre el sistema de IA y el entorno en el que opera.

En la línea de lo señalado, el Reglamento (UE) de Inteligencia Artificial (RIA)[4] promueve un enfoque europeo de la IA centrado en el riesgo y tiene por objeto garantizar la protección de los derechos fundamentales y la seguridad de los usuarios en los ámbitos de especial incidencia. Uno de los ámbitos de alto riesgo predefinidos en el RIA (Anexo III) es la relación laboral en el que uno de los usuarios importantes o grupos de usuarios que se espera que interactúen o aprovechen la IA, es la empresa y, a la vez, el trabajador, que desarrolla su actividad laboral expuesto, no solamente a los riesgos específicos del puesto de trabajo, sino a los derivados de la IA.

El acercamiento al problema de los riesgos laborales de la IA no puede ser realizado, basándose simplemente en los enunciados de la normativa de la IA, porque en realidad el RIA no se ocupa de este tema. Aborda los riesgos de la IA de manera genérica, prohibiendo o limitando el uso de sistemas de IA que presenten un riesgo inaceptable para la seguridad, la salud, la dignidad o la autonomía de las personas, o que violen los valores democráticos. No obstante, las normas armonizadas que se establecen en el RIA deben entenderse -según indica el Considerando 9- sin perjuicio del Derecho vigente de la Unión, en particular en materia de protección de datos, protección de los consumidores, derechos fundamentales, empleo, protección de los trabajadores y seguridad de los productos, al que complementa el presente Reglamento.

La aproximación al tema preventivo laboral del RIA requiere, por tanto, un enfoque conjunto, en el que se tome en consideración el orden jurídico laboral, en particular,

[4] El día 12 de julio de 2024 el RIA se publicó en el Diario Oficial de la Unión Europea 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.o 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial o AI Act), en lo sucesivo RIA o el Reglamento simplemente), que entró en vigor el 1 de agosto de 2024, aunque su aplicación se producirá de forma diferida. Será aplicable a partir del 2 de agosto de 2026. No obstante: a) los capítulos I y II serán aplicables a partir del 2 de febrero de 2025; b) el capítulo III, sección 4, el capítulo V, el capítulo VII y el capítulo XII y el artículo 78 serán aplicables a partir del 2 de agosto de 2025, a excepción del artículo 101; c) el artículo 6, apartado 1, y las obligaciones correspondientes del presente Reglamento serán aplicables a partir del 2 de agosto de 2027.

el marco preventivo laboral, porque, como se advierte en el Considerando 9, el RIA no debe afectar al Derecho de la Unión en materia de política social ni a la legislación laboral nacional —conforme al Derecho de la Unión— relativa a las condiciones de empleo y de trabajo, incluidas la salud y seguridad en el trabajo y la relación entre empleadores y trabajadores.

Ahora bien, en seguida surgen las fricciones porque la conceptuación del riesgo de la normativa de IA y la de la normativa de prevención de riesgos laborales divergen notablemente. Como se verá, la regulación de la IA atiende a la magnitud del riesgo y a criterios extrasistemáticos de matriz económica, pues, aparte de prever una clasificación de los sistemas de IA según el nivel de riesgo, busca promover la innovación y la competitividad en el sector de la IA, procurando no obstaculizar el desarrollo económico[5]. En cambio, el concepto de riesgo recabado de la normativa de seguridad y salud laboral no establece categorización alguna y se basa únicamente en la protección de vida y la seguridad y salud de la persona del trabajador.

Aquí se trata de conectar ambas soluciones valorativas, de integrar el complejo entramado de principios de los dos ordenamientos jurídicos concurrentes, y de extraer criterios racionales para alcanzar un punto de equilibrio entre los dos sistemas de protección de riesgo, viendo qué obligaciones deben observar los proveedores y cuáles los desarrolladores de los sistemas de IA (empresarios) a partir de los requisitos del RIA y los específicos de prevención de los riesgos laborales.

## 2. La noción de riesgo protegido en la LPRL

Antes de adentrarnos en la conceptualización del riesgo en el RIA, procede exponer sucintamente el significado y alcance de riesgo laboral y su prevención en el ámbito del ordenamiento jurídico laboral, pues, al contrario de lo que somos inducidos a creer por la modernidad de la materia regulada, no se antepone el marco normativo regulador de la IA por ser la norma posterior, sino que, al contrario debe prevalecer la normativa de prevención laboral porque, como se ha señalado, la nueva normativa de protección en materia de IA viene a complementar la normativa de prevención de riesgos laborales y no al revés.

Es importante subrayar esto, ya que la normativa laboral preventiva no debe doblegarse al servicio de los objetivos de la IA, sino que ésta debe interpretarse a la luz de las irreductibles especificidades de aquella. Y ello porque la normativa de IA ha de tratar

[5] Considerando 8: "*En consecuencia, se necesita un marco jurídico de la Unión que establezca unas normas armonizadas en materia de IA para impulsar el desarrollo, la utilización y la adopción en el mercado interior de la IA y que, al mismo tiempo, ofrezca un nivel elevado de protección de los intereses públicos, como la salud y la seguridad y la protección de los derechos fundamentales, en particular la democracia, el Estado de Derecho y la protección del medio ambiente, reconocidos y protegidos por el Derecho de la Unión. Para alcanzar dicho objetivo, conviene establecer normas que regulen la introducción en el mercado, la puesta en servicio y la utilización de determinados sistemas de IA, lo que garantizará el buen funcionamiento del mercado interior y permitirá que dichos sistemas se beneficien del principio de libre circulación de mercancías y servicios*".

de conectarse con otros modelos de regulación vigente en esta materia o en otras como la de seguridad del producto.

Como pone de manifiesto el Considerando 64, los peligros de los sistemas de IA abarcados por los requisitos del Reglamento IA "*se refieren a aspectos diferentes de los contemplados en los actos de armonización de la Unión existentes*" y, por consiguiente, los requisitos del Reglamento IA "*completa*(n) *el conjunto existente de actos de armonización de la Unión*". Por ejemplo, se señala que las máquinas que incorporan un sistema de IA pueden presentar riesgos de los que no se ocupan los requisitos esenciales de salud y seguridad establecidos en la legislación armonizada pertinente de la Unión, ya que esa legislación sectorial no aborda los riesgos específicos de los sistemas de IA.

Así las cosas, es preciso conocer el contexto en el que debe ser aplicada la nueva normativa europea en materia de IA. En este sentido, cabe recordar, aunque sea un lugar común, que la normativa general de prevención de riesgos laborales (la Directiva 89/391//CEE, relativa a la aplicación de medidas para promover la mejora de la seguridad y salud de los trabajadores), y la norma de trasposición a nuestro ordenamiento (la LPRL), imponen un deber empresarial de garantizar la seguridad y salud de los trabajadores frente a los riesgos laborales.

La normativa de prevención de riesgos laborales concibe el "riesgo laboral" como *la posibilidad de que un trabajador sufra un determinado daño derivado del trabajo* (art. 4.2 LPRL). Dicha normativa tiene por objeto identificar y estimar la magnitud del riesgo para posteriormente evitarlo o, en su caso, reducirlo y controlarlo. Entre los principios generales que integran el deber general de prevención laboral (art. 15.1 LPRL), se hallan los de: a) evitar los riesgos; b) evaluar los riesgos que no se puedan evitar; c) combatir los riesgos en su origen.

La LPRL establece un criterio general de protección de la seguridad y salud laboral del trabajador, sin excluir, ni distinguir ningún tipo de riesgo. Considera que, para calificar un riesgo desde el punto de vista de su gravedad, se debe atender conjuntamente a la probabilidad de que se produzca el daño y a la severidad del mismo (art. 4.2 LPRL).

El empresario contrae con el trabajador una deuda de seguridad, que no se satisface con el mero cumplimiento formal de la normativa en materia de prevención de riesgos laborales. El empresario está obligado a adoptar cuantas medidas sean necesarias para evitar el daño, desarrollando una acción permanente de seguimiento de la actividad preventiva, identificando, evaluando y controlando los riesgos que no se hayan podido evitar, y de los niveles de protección (art. 14.2 LPRL).

La deuda de seguridad requiere una diligencia constante por parte del empleador para evitar que el daño se produzca. Incumbe al empleador proteger al trabajador incluso frente a sus propias imprudencias profesionales. Por ello, entre las medidas preventivas que debe adoptar se incluye la de prever las distracciones o imprudencias no temerarias que el empleador pudiera cometer en el desarrollo de su actividad laboral (art. 15.4 LPRL).

El empresario, para evitar la responsabilidad, ha de acreditar haber agotado toda diligencia posible incluso más allá de las exigencias reglamentarias. Solo queda exonerado de responsabilidad si el resultado lesivo se hubiese producido por fuerza mayor o caso

fortuito, por negligencia exclusiva no previsible del trabajador o por culpa exclusiva de terceros no evitable por el empresario. Aun y todo, corresponde al empresario acreditar la concurrencia de esa posible causa de exoneración, en tanto que él es el titular de la deuda de seguridad y habida cuenta de los términos cuasiobjetivos en que la misma está concebida legalmente [STS (Sala 4ª) 30 de junio de 2010, R. 4123/2008].

## 3. La noción de riesgo en el RIA

La normativa de IA comparte elementos con la normativa de prevención de riesgos laborales: por un lado, sitúa el centro de gravedad en la valoración del riesgo[6], que es inherente a cualquier actividad social o económica; y, por otro, persigue como objetivo controlar el riesgo.

En efecto, tal como se ha señalado, el RIA ha sido diseñado con un enfoque basado en el riesgo. El sistema de protección frente a los perjuicios que conlleva el uso de los medios y sistemas de IA se construye sobre la noción de riesgo para los derechos, la seguridad y el buen funcionamiento del mercado único del Unión Europea. El riesgo se define como "*la combinación de la probabilidad de que se produzca un daño y la gravedad del daño*" (art. 3.2 RIA). De forma que, el concepto de daño o perjuicio se erige en la piedra angular sobre el que se asienta el riesgo.

Por otra parte, el RIA persigue el mismo objeto que la LPRL, porque trata de identificar y estimar la magnitud del riesgo para posteriormente evitarlo o, en su caso, reducirlo y controlarlo. El RIA impone a los proveedores obligaciones orientadas a evaluar riesgos concretos y a aplicar medidas de reducción del riesgo razonable. Su control constituye, por tanto, el objetivo clave de la actividad preventiva.

No obstante, ambas legislaciones presentan, algunas diferencias importantes:

### *3.1. La categorización del riesgo en el RIA: distinción entre riesgo sistémico y riesgo crónico*

Difieren sobre la categorización de riesgo. La LPRL establece un criterio general, sin excluir, ni distinguir ningún tipo de riesgo. Considera, como se ha observado, que, para calificar un riesgo desde el punto de vista de su gravedad, se debe atender conjuntamente a la probabilidad de que se produzca el daño y a la severidad del mismo (art. 4.2 LPRL).

Sin embargo, de acuerdo con la propia configuración del RIA, es preciso distinguir entre dos grandes categorías: por un lado, los riesgos extremos o sistémicos, seguramente de muy reducida probabilidad en el ámbito laboral, pero de una intensidad e implicaciones mucho más graves; y, por otro lado, los riesgos crónicos de alta frecuencia pero de intensidad moderada.

[6] MERCADER UGUINA, J.: "Los usos de alto riesgo en el ámbito laboral de la IA y la certificación", *El Foro de Labos, 9/5/2024*, Disponible en: https://www.elforodelabos.es/2024/05/los-usos-de-alto-riesgo-en-el-ambito-laboral-de-la-ia-y-la-autocertificacion/

Inicialmente la Propuesta de Reglamento de IA solo contemplaba los riesgos crónicos, pero en su proceso normativo de elaboración, especialmente su fase final, a consecuencia de la aparición de los denominados GPAI, (los General Purpose AI sistems and models) se ha insertado una regulación de los modelos y sistemas de IA de uso general[7].

Los riegos extremos abarcarían los modelos de sistemas de IA general, que, por el incremento de capacidades y la autonomía de estos sistemas, "podrían amplificar el impacto de la IA, planteando unos riesgos que incluyen daños sociales a gran escala, usos maliciosos y una pérdida irreversible del control humano sobre los sistemas autónomos de IA"[8].

Se incluiría en esta categoría el llamado «riesgo sistémico», que el art. 3. 65 RIA, describe como "*un riesgo específico de las capacidades de gran impacto de los modelos de IA de uso general, que tienen unas repercusiones considerables en el mercado de la Unión debido a su alcance o a los efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto, que puede propagarse a gran escala a lo largo de toda la cadena de valor*"[9]. Este riesgo lo presentan aquellos productos elaborados con IA de aparición no tan frecuente en el lugar de trabajo, pero de posibles efectos devastadores para la seguridad y salud de las personas e integridad de los bienes.

Un modelo de IA de uso general se considerará que es de riesgo sistémico si reúne alguna de las siguientes condiciones:

a) tener capacidades de gran impacto evaluadas a partir de herramientas y metodologías técnicas adecuadas, como indicadores y parámetros de referencia: se presumirá que un modelo de IA de uso general tiene las referidas capacidades de gran impacto cuando la cantidad acumulada de cálculo utilizada para su entrenamiento, medida en operaciones de coma flotante, sea superior a 1025) (art. 51.2 RIA).

---

[7] BARRIO ANDRÉS, M. : "Algunos claroscuros en el Reglamento Europeo de Inteligencia Artificial", *Diario LA LEY, Nº 86, Sección Ciberderecho*, 30 de Julio de 2024.

[8] Parafraseando al grupo de expertos del documento AA. VV.: "Managing extreme AI risks amid rapid progress", *Science* , 20 may 2024, Vol. 384: https://www.science.org/doi/10.1126/science.adn0117

[9] Considerando 110: "*En particular, los enfoques internacionales han establecido hasta la fecha la necesidad de prestar atención a los riesgos derivados de posibles usos indebidos intencionados o de problemas en materia de control relacionados con la armonización con la intención humana no deseados, a los riesgos químicos, biológicos, radiológicos y nucleares, como las maneras en que las barreras a la entrada pueden reducirse, en particular para el desarrollo, el diseño, la adquisición o el uso de armas, a las cibercapacidades ofensivas, como las maneras en que pueden propiciarse el descubrimiento, la explotación o el uso operativo de vulnerabilidades, a los efectos de la interacción y el uso de herramientas, incluida, por ejemplo, la capacidad de controlar sistemas físicos e interferir en el funcionamiento de infraestructuras críticas, a los riesgos derivados del hecho que los modelos hagan copias de sí mismos o se «autorrepliquen» o entrenen a otros modelos, a las maneras en que los modelos pueden dar lugar a sesgos dañinos y discriminación que entrañan riesgos para las personas, las comunidades o las sociedades, a la facilitación de la desinformación o el menoscabo de la intimidad, que suponen una amenaza para los valores democráticos y los derechos humanos, al riesgo de que un acontecimiento concreto dé lugar a una reacción en cadena con efectos negativos considerables que podrían afectar incluso a una ciudad entera, un ámbito de actividad entero o una comunidad entera*".

b) tener, con arreglo a una decisión de la Comisión, adoptada de oficio o a raíz de una alerta cualificada del grupo de expertos científicos, las capacidades o un impacto equivalente a los establecidos en la letra a), de acuerdo con los criterios establecidos en el anexo XIII[10] (art. 51.1 RIA).

No obstante, la Comisión Europea puede modificar los referidos umbrales para clasificar los modelos de IA de uso general —los modelos GPAI— como de riesgo «sistémico» en función de los avances tecnológicos, como las mejoras algorítmicas o la mayor eficiencia del hardware, cuando sea necesario, para que los umbrales reflejen el estado actual de la técnica (arts. 51.3 y 52.4 RIA). En consecuencia, la Comisión tiene la oportunidad de utilizar pruebas del mundo real para establecer y definir el umbral de riesgo sistémico yendo más allá de los FLOP y añadiéndolos o sustituyéndolos por nuevos criterios de referencia[11].

Dentro de la segunda categoría llamada de "riesgos crónicos" cabe considerar los sistemas de IA del Anexo III de efectos mucho más moderados. En el marco conceptual del RIA, el riesgo crónico se construye sobre una escala de riesgos en función de la capacidad para inferir daños para las personas y la sociedad y vulneraciones a derechos fundamentales. Se trata de un riesgo relacionado con la puesta en marcha y uso de determinados sistemas de IA que deberán cumplir los estándares del RIA para poder operar en Europa y de esta manera, quizá de forma indirecta, proteger a los ciudadanos europeos[12].

---

[10] ANEXO XIII: "*Criterios para la clasificación de los modelos de IA de uso general con riesgo sistémico a que se refiere el artículo 51:*

*Con el fin de determinar si un modelo de IA de uso general tiene unas capacidades o unos efectos equivalentes a los contemplados en el artículo 51, apartado 1, letra a), la Comisión tendrá en cuenta los siguientes criterios:*

*a) el número de parámetros del modelo;*

*b) la calidad o el tamaño del conjunto de datos, por ejemplo medidos a través de criptofichas;*

*c) la cantidad de cálculo utilizada para entrenar el modelo, medida en operaciones de coma flotante o indicada con una combinación de otras variables, como el coste estimado del entrenamiento, el tiempo estimado necesario o el consumo de energía estimado para el mismo;*

*d) las modalidades de entrada y salida del modelo, como la conversión de texto a texto (grandes modelos de lenguaje), la conversión de texto a imagen, la multimodalidad y los umbrales punteros para determinar las capacidades de gran impacto de cada modalidad, y el tipo concreto de entradas y salidas (por ejemplo, secuencias biológicas);*

*e) los parámetros de referencia y las evaluaciones de las capacidades del modelo, también teniendo en cuenta el número de tareas sin entrenamiento adicional, la adaptabilidad para aprender tareas nuevas distintas, su nivel de autonomía y capacidad de ampliación y las herramientas a las que tiene acceso;*

*f) si sus repercusiones para el mercado interior son importantes debido a su alcance, lo que se dará por supuesto cuando se haya puesto a disposición de al menos 10 000 usuarios profesionales registrados establecidos en la Unión;*

*g) el número de usuarios finales registrados*".

[11] BARRIO ANDRÉS, M. : "Algunos claroscuros en el Reglamento Europeo de Inteligencia Artificial", op cit.

[12] VESTRI, G.: "La Unión Europea estrena el Reglamento de Inteligencia Artificial (RIA). Control, supervisión y uso de una tecnología cada vez más presente en la vida de todos", op. cit.

### *3.2. La pirámide de riesgos crónicos*

En el llamado riesgo crónico, el RIA establece una clasificación, dando lugar a lo que venimos calificando como una "pirámide de riesgos"[13] integrado por cuatro diferentes niveles de riesgo: inaceptable, alto, limitado y mínimo. En la base se sitúa el riesgo mínimo por su menor capacidad para generar daños, y en el nivel más alto el riesgo inaceptable o prohibido por ser contrario a los valores de la Unión. Y en sendos escalones intermedios se sitúan los sistemas de riesgo alto y los de riesgo medio.

El RIA incorpora, en el artículo 5, una serie de prácticas sobre las que extenderá un total veto a su introducción en el mercado, puesta en servicio o utilización. Se consideran como tales los sistemas de IA que integren técnicas deliberadamente manipuladoras o engañosas, así como técnicas subliminales que trasciendan la consciencia [art. 5.1. a) RIA]; aquellos que exploten "alguna de las vulnerabilidades de una persona o un grupo específico de personas derivadas de su edad o discapacidad, o de una situación social o económica específica, con el objetivo o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho grupo de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona o a otra" [art. 5.1 b) RIA]; sistemas de IA que permitan "evaluar o clasificar a personas físicas o a colectivos de personas durante un período determinado de tiempo atendiendo a su comportamiento social o a características personales o de su personalidad conocidas, inferidas o predichas" ( esto es, la asignación de la conocida como "puntuación social" ("social scoring") por parte de las autoridades públicas) [art. 5.1.c) RIA]; "sistemas de IA que creen o amplíen bases de datos de reconocimiento facial mediante la extracción no selectiva de imágenes faciales de internet o de circuitos cerrados de televisión" [art. 5.1.e) RIA]; "sistemas de IA para inferir las emociones de una persona física en los lugares de trabajo y en los centros educativos, excepto cuando el sistema de IA esté destinado a ser instalado o introducido en el mercado por motivos médicos o de seguridad" [art. 5.1.f) RIA]; "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]; "sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público con fines de garantía del cumplimiento del Derecho, salvo y en la medida en que dicho uso sea estrictamente necesario" [art. 5.1.h) RIA].

En un segundo nivel, el RIA sitúa los sistemas calificados de alto riesgo, regulados en el art. 6 y ss, que "constituyen el verdadero núcleo principal de la regulación europea de la IA", a los que dedica la mayor parte de su contenido. Los sistemas de IA, que el RIA califica

[13] GOÑI SEIN, J. L.: "El Reglamento UE de Inteligencia Artificial y su interrelación con la normativa de seguridad y salud en el trabajo", AA. VV. (Dirs. EGUSQUIZA, M. A.; RODRÍGUEZ SANZ DE GALDEANO, B.): *Inteligencia artificial y prevención de riesgos laborales: obligaciones y responsabilidades*, Tirant lo Blanch, Valencia 2023, p. 83.

de alto riesgo[14], se definen por dos rasgos (Considerando 46 y art. 6.1 y 2 RIA): 1) porque "*la introducción en el mercado de la Unión, la puesta en servicio o la utilización de sistemas de IA de alto riesgo debe supeditarse al cumplimiento de determinados requisitos obligatorios* antes de su comercialización (sistema de gestión de riesgos, calidad de datos, transparencia, supervisión humana); y 2) porque sus efectos perjudiciales no deben entrañar "*riesgos inaceptables para intereses públicos importantes de la UE, reconocidos y protegidos por el Derecho de la Unión*".

Se reconoce abiertamente que los sistemas de IA de alto riesgo " *pueden tener un efecto adverso para la salud y la seguridad de las personas, en particular cuando funcionan como componentes de seguridad de productos*" (Considerando 47), es decir, que se caracterizan por conllevar un grado importante de efectos perjudiciales, pero sin llegar al grado de "inaceptables"; o sea, sin alcanzar la magnitud de las consecuencias adversas de las prácticas prohibidas en el art. 5 RIA.

En el marco conceptual del RIA, el alto riesgo se refiere a unos ámbitos predefinidos especificados en el Anexo III del Reglamento de IA. No obstante, la Comisión Europea está facultada para añadir, modificar o suprimir los casos de uso de los sistemas de IA de alto riesgo del Anexo III (art. 7, apartados 1 y 3, RIA) y para modificar o añadir nuevas condiciones en las que los sistemas de IA de alto riesgo del Anexo III no se considerarán de alto riesgo con arreglo al artículo 6.3 del RIA (art. 6, apartados 6 y 7, RIA). Al llevarlo a cabo, la Comisión debe sopesar explícitamente los beneficios económicos y sociales de los sistemas de IA frente a los riesgos, basándose en pruebas empíricas suficientes[15].

En un tercer nivel el RIA contempla los sistemas de IA que representan un Riesgo limitado. Son sistemas de IA que pueden influir en los Derechos o la voluntad de los usuarios, pero en menor medida que los sistemas de alto riesgo. Estos sistemas están sujetos a requisitos de transparencia o información. En concreto: en los sistemas de IA diseñados para interactuar con personas (chatbots) se debe informar de que se está tratando con un sistema de IA; en los sistemas de IA, incluidos los sistemas de uso general, se garantizará que los resultados del sistema de IA estén marcados en un formato legible por máquina y detectable como generado o manipulado artificialmente; en sistemas de reconocimiento de emociones y sistemas de clasificación biométrica se deberá informar de que se están usando; en los sistemas que generen o manipulen imágenes o audio de personas (ultrafalsificación) se deberá informar de que se trata de imágenes o sonidos

[14] Según el art. 6.1 RIA, *un sistema de IA se considerará de alto riesgo cuando reúna las dos condiciones que se indican a continuación:*

*a) que el sistema de IA esté destinado a ser utilizado como componente de seguridad de un producto que entre en el ámbito de aplicación de los actos legislativos de armonización de la Unión enumerados en el anexo I, o que el propio sistema de IA sea uno de dichos productos, y*

*b) que el producto del que el sistema de IA sea componente de seguridad con arreglo a la letra a), o el propio sistema de IA como producto, deba someterse a una evaluación de la conformidad de terceros para su introducción en el mercado o puesta en servicio con arreglo a los actos legislativos de armonización de la Unión enumerados en el anexo I.*

*2. Además de los sistemas de IA de alto riesgo a que se refiere el apartado 1, también se considerarán de alto riesgo sistemas de IA contemplados en el anexo III.*

[15] BARRIO ANDRÉS, M. : "Algunos claroscuros en el Reglamento Europeo de Inteligencia Artificial", op cit.

manipulados artificialmente. Existe el derecho a saber que se está hablando con un bot (en lugar de un humano) y que una imagen es creada o modificada por IA.

En último lugar, se menciona el Riesgo mínimo o bajo, que integra a los sistemas de IA que no tienen impacto directo en los Derechos fundamentales o la seguridad de las personas, y que ofrecen amplias opciones y control a los usuarios. Estos sistemas están libres de cualquier obligación normativa, para fomentar la innovación y la experimentación. Se alude, en concreto, a los sistemas de IA utilizados para fines lúdicos (como videojuegos) o puramente estéticos (como filtros fotográficos)[16]. Los fabricantes pueden acogerse voluntariamente al RIA, aplicando alguno o todos los requisitos establecidos en el capítulo III, sección 2 para los sistemas de alto riesgo, o adherirse a códigos de conducta para una IA confiable, que los contemplen, elaborados por ellos o por las organizaciones a las que pertenecen (art. 95 RIA).

### *3.3. Ámbitos de riesgo crónico predefinidos con relación al ámbito laboral*

La clasificación de los sistemas de IA según el nivel de riesgo, presenta algunas prácticas que tienen su ámbito de proyección natural en la relación laboral.

En las prácticas prohibidas, como ya se ha señalado, queda proscrito el "uso de sistemas de IA para inferir las emociones[17] de una persona física en los lugares de trabajo (…), excepto cuando el sistema de IA esté destinado a ser instalado o introducido en el mercado por motivos médicos o de seguridad [art. 5.1.f) RIA]. De este modo, "quedan vetadas las cada vez más frecuentes técnicas o métodos de análisis de aptitudes, habilidades y capacidades psicosociales de los procesos selectivos de las empresas cuando se integren, por ejemplo, valiéndose de las destrezas de los nuevos softwares utilizados en las entrevistas virtuales, sistemas para inferir las emociones y crear con ello un perfil psicotécnico de los candidatos, o detectar el estado emocional de éste a lo largo de la entrevista"[18].

En este sentido, caerían dentro de lo ilícito "los sistemas de IA integrados en *wereables* o plataformas de gestión laboral que procuren inferir información sobre estos estados

[16] VESTRI, G.: "La Unión Europea estrena el Reglamento de Inteligencia Artificial (RIA). Control, supervisión y uso de una tecnología cada vez más presente en la vida de todos", op. cit

[17] Según el Considerando 18, "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. Esas expresiones pueden ser expresiones faciales básicas, como un ceño fruncido o una sonrisa; gestos como el movimiento de las manos, los brazos o la cabeza, o características de la voz de una persona, como una voz alzada o un susurro".

[18] GOÑI SEIN,J.L.; RODRÍGUEZ SANZ DE GALDEANO, B.; LLORENS ESPADA, J.; MARIN MALO, M.: "El impacto del nuevo marco normativo europeo de la inteligencia artificial en las relaciones laborales", AA VV (Dir. RICHARD GONZÁLEZ, M.), en prensa edit J B. Bosch,

de ánimo o emociones de los trabajadores, valiéndose de cualquier medio como pudieran ser sensores biométricos, señales fisiológicas tales como la temperatura corporal, frecuencia cardíaca, resistencia de la piel y onda del pulso , reconocimiento facial , voz o incluso el procesamiento de datos que incluyan conductas del trabajador de las que pueda inferirse el estado de ánimo, como por ejemplo el uso concreto que se hace del smartphone"[19].

En cambio, sí podrán tener cabida aquellos dispositivos que, habiendo sido previstos como una medida integrada en el Plan de prevención de riesgos laborales, busquen detectar el estado físico del trabajador con el objeto de prevenir futuros accidentes de trabajo, o su utilización responda a objetivos médicos. Esto genera que estas técnicas deban someterse al bloque normativo de vigilancia de la salud laboral ex art. 22 LPRL como cualquier otro reconocimiento médico de salud laboral[20].

Con relación a los sistemas de alto riesgo (art. 6.2 RIA), las actividades de IA que tienen una directa proyección en el ámbito laboral, se encuentran especificados en el punto 4 del Anexo III. En concreto, se consideran como de alto riesgo los que afecten al "*empleo, gestión de los trabajadores y acceso al autoempleo*", en particular, 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 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 la personas en el marco de dichas relaciones*".

Se comprueba fácilmente que el Anexo III no constituye una relación acabada de potenciales ámbitos de alto riesgo derivados de sistemas de IA utilizados en el ámbito laboral, sino una relación más bien indicativa. En el ámbito laboral predefinido por el RIA, la mayor parte de actividades enumeradas constituyen gestión de recursos humanos. Pero puede haber también otros sistemas de IA no catalogados como de alto riesgo que podrían tener dicha consideración en atención a los riesgos que entrañan, como, por ejemplo, los sistemas de IA incorporados a máquinas con funciones de eficiencia o seguridad de bienes y personas, o, en materia preventiva, los equipos de protección individual inteligentes, o las plataformas digitales para la gestión de la PRL, y en general cualquier plataforma 4.0 que integre "innovaciones tecnológicas, como sistemas cognitivos, que son implementados a través de la aplicación de la Inteligencia Artificial a los datos, redes neuronales convolucionales (CNN) y aprendizaje por refuerzo profundo (DRL), haciendo que sean capaces de controlar un enorme conjunto de parámetros relacionados con procesos y el entorno"[21].

---

[19] Idem

[20] Idem

[21] LLORENS ESPADA, J.: "Inteligencia artificial y salud laboral", AA. VV. (Dirs. EGUSQUIZA, M. A.; RODRÍGUEZ SANZ DE GALDEANO, B.): *Inteligencia artificial y prevención de riesgos laborales: obligaciones y responsabilidades*, Tirant lo Blanch, Valencia 2023, p. 221.

Desde la perspectiva del marco normativo de prevención de riesgos laborales, conviene tener presente la distinta valoración de la relevancia del riesgo. A diferencia del RIA que fragmenta y limita el riesgo tomado en consideración, en la LPRL el "riesgo laboral" se concibe de manera íntegra y unitaria como la posibilidad de que un trabajador sufra un determinado daño derivado del trabajo (art. 4.2 LPRL) sin impropias distinciones. Lo que obliga al empresario a prever cualquier riesgo, en cualquier ámbito que se produzca, y con independencia de la gravedad.

De ahí que, el RIA advierta, en el Considerando (63), que "(e)l *hecho de que un sistema de IA sea clasificado como un sistema de IA de alto riesgo en virtud del presente Reglamento no debe interpretarse como indicador de que su uso sea legal con arreglo a otros actos del Derecho de la Unión o del Derecho nacional compatible con el Derecho de la Unión*", añadiendo que todo "*uso de ese tipo debe seguir realizándose exclusivamente en consonancia con los requisitos oportunos derivados de la Carta y de los actos aplicables del Derecho derivado de la Unión y del Derecho nacional*".

Ello nos lleva a afirmar, en línea de lo observado anteriormente, que la regulación de los sistemas de riesgo alto no se agota en sus propios requisitos, sino que conlleva una ulterior tarea de aplicación de los requisitos de Derecho derivado de la Unión y del Derecho nacional, en particular, respecto de la materia de prevención laboral, la Directiva 89/391 CEE, Directiva marco sobre salud y seguridad en el trabajo, de 12 de junio de 1989[22], y la LPRL.

A ello debe añadirse, además, que la vinculación del RIA exclusivamente a los sistemas de IA clasificados como de alto riesgo, no significa que, respecto del resto de los sistemas de IA no calificados como tal, no rija la normativa de prevención de riesgos laborales y, por tanto, no se aplique el deber del empresario de protección de los trabajadores frente a los posibles riesgos. La obligación del empresario de garantizar la seguridad y salud laboral, y de extremar la vigilancia en el cumplimiento de las normas de seguridad, es un principio de las relaciones laborales, que se aplica por igual a cualquier sistema de IA que entrañe cualquier tipo de riesgo laboral.

### *3.4. La valoración del riesgo: criterios de matriz económica*

En el RIA se observa, aparte, un segundo elemento diferenciador con respecto a la LPRL, pues el RIA ha incorporado criterios extrasistemáticos de matriz económica, en la determinación del riesgo protegido. Conviene no olvidar que el RIA se propone crear un mercado único para la IA, facilitando la libre circulación y el reconocimiento de los sistemas de IA que cumplan con las normas de la UE. Lo cual obliga al intérprete a tener en cuenta inevitablemente en el horizonte hermenéutico del RIA valoraciones de política

[22] Pese a no estar incluida en la "Lista de actos legislativos de armonización de la Unión" del Anexo I, que en esta materia, tan solo menciona el Reglamento (UE) 2016/425 del Parlamento Europeo y del Consejo, de 9 de marzo de 2016, relativo a los equipos de protección individual y por el que se deroga la Directiva 89/686/CEE del Consejo (DO L 81 de 31.3.2016, p. 51)

de derecho, señaladamente, económicas, de comercio internacional y de funcionamiento de mercado

Esas consideraciones de orden económico no se aplican a los modelos de IA de uso general que pueden plantear los riesgos sistémicos, provocando, por ejemplo, "*cualquier efecto negativo real o razonablemente previsible en relación con accidentes graves, perturbaciones de sectores críticos y consecuencias graves para la salud y la seguridad públicas, cualquier efecto negativo real o razonablemente previsible sobre los procesos democráticos y la seguridad pública y económica o la difusión de contenidos ilícitos, falsos o discriminatorios*" (Considerando 110), sino a los riesgos crónicos o menores.

Con respecto a la categoría de alto riesgo del Anexo III, de efectos mucho más moderados, el RIA incorpora matices importantes y algunos criterios limitativos de política de derecho que le alejan mucho, del ordenamiento jurídico laboral preventivo.

El riesgo tomado en consideración se contrae, por lo pronto, a los sistemas de IA que presentan un "riesgo considerable" de ser perjudiciales para la salud y la seguridad o los derechos fundamentales de las personas, "teniendo en cuenta tanto la gravedad del posible perjuicio como la probabilidad de que se produzca" (Considerando 52). En este sentido, precisa el Considerando 53, que "es importante aclarar que pueden existir casos específicos en los que los sistemas de IA relativos a ámbitos predefinidos especificados en el presente Reglamento no entrañen un riesgo considerable de causar un perjuicio a los intereses jurídicos amparados por dichos ámbitos, dado que no influyen sustancialmente en la toma de decisiones o no perjudican dichos intereses sustancialmente". Es decir que, dentro de estos últimos, el sistema de obligaciones y garantías de RIA se extiende solo a aquellos que entrañen un riesgo considerable de causar un perjuicio a los intereses jurídicos amparados por dichos ámbitos, excluyendo a los que no influyen sustancialmente en la toma de decisiones o no perjudican dichos intereses sustancialmente (Considerando 53).

Pero ahí no acaba todo, porque el RIA restringe, aun más, los sistemas de IA sujetos al régimen obligacional del RIA, al obligar a tener en cuenta, además, valoraciones de política de derecho, y, más en concreto, valoraciones económicas de comercio internacional. Abiertamente se dispone en el Considerando (46) que: "*A fin de garantizar la coherencia y evitar una carga administrativa innecesaria o costes innecesarios, los proveedores de un producto que contenga uno o varios sistemas de IA de alto riesgo, a los que se apliquen los requisitos del presente Reglamento o de los actos legislativos de armonización de la Unión enumerados en un anexo del presente Reglamento, deben ser flexibles en lo que respecta a las decisiones operativas relativas a la manera de garantizar la conformidad de un producto que contenga uno o varios sistemas de IA con todos los requisitos aplicables de la legislación de armonización de la Unión de manera óptima. La clasificación de un sistema de IA como «de alto riesgo» debe limitarse a aquellos sistemas de IA que tengan un efecto perjudicial importante en 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*".

A la vista de lo cual, el intérprete tiene que introducir inevitablemente en el horizonte hermenéutico del RIA valoraciones de política de derecho, señaladamente, económicas, de comercio internacional y de funcionamiento de mercado. Criterios que obvia-

mente el interprete de la normativa de prevención laboral debe excluir por considerarlos ajenos a los únicos intereses en juego de la preservación de la seguridad y salud laboral del trabajador.

Parece que la "calificación 'de alto riesgo' "se limita a aquellos sistemas de IA que tengan consecuencias perjudiciales importantes para la salud, la seguridad y los derechos fundamentales de las personas de la Unión", y debiendo además, reducirse "al mínimo cualquier posible restricción del comercio internacional, si la hubiera". Siendo así, se adopta una idea de alto riesgo bastante restrictiva, donde el objetivo de la seguridad y salud se subordina de alguna manera a las consideraciones económicas de no imponer restricciones innecesarias al comercio y de comprometer el desarrollo del mercado único.

En suma, la calificación de alto riesgo del sistema de IA dependerá del contexto de su específico de acuerdo con los criterios establecidos en el art. 6.3 RIA y los elementos interpretativos expuestos en los Considerandos 52 y 53 del RIA[23].

Ello tendrá, como ya he comentado en un anterior trabajo[24], un doble efecto significativo sobre la aplicación de los sistemas de IA en el lugar de trabajo, quedando probablemente excluidos buena parte de aquellos sistemas de IA. En primer lugar, porque no representen un efecto nocivo grave o alto sobre los trabajadores; de forma que, respecto de estos sistemas de IA, aunque representen un peligro para los trabajadores, no será necesario el cumplimiento de los requisitos esenciales de alto riesgo y simplemente se exigirán obligaciones específicas de trasparencia de los riesgos limitados o mínimos, como, por ejemplo, que los usuarios sean conscientes de que están interactuando con una máquina.

Y, en segundo lugar, porque en la mayor parte de los sistemas de IA considerados de alto riesgo en el trabajo, los efectos apreciables de impacto negativo en la seguridad de las personas son de carácter psicológico (por ej. el estrés y patologías psicosomáticas derivadas del monitoreo continuo de la actividad del trabajador o la conectividad constante en el trabajo en plataformas) y se van generando paulatinamente, de forma que, a priori podría no ser considerado de alto riesgo a la luz de la Ley de IA, porque, en realidad, como observan KULLMAN y CEFALIELLO, el impacto nocivo no aparece inmediatamente; es un proceso gradual"[25]

No obstante, aquellos proveedores que consideren que un sistema de IA no es de alto riesgo, pese a estar contemplado en el anexo III, deberán realizar y documentar una evaluación antes de que dicho sistema sea introducido en el mercado o puesto en servicio (art. 6.4 RIA). En esta evaluación se debe acreditar que el sistema de IA no plantea

[23] FERNÁNDEZ HERNÁNDEZ, C. Y EGUILUZ CASTAÑEIRA, J. A.: "Diez puntos críticos del Reglamento europeo de Inteligencia Artificial", *Diario La Ley*, nº 85, *Sección Ciberderecho*, 28 de junio de 2024.

[24] GOÑI SEIN, J. L.: "El Reglamento UE de Inteligencia Artificial y su interrelación con la normativa de seguridad y salud en el trabajo", AA. VV. (Dirs. EGUSQUIZA, M. A.; RODRÍGUEZ SANZ DE GALDEANO, B.): *Inteligencia artificial…,* op. cit. pp. 103-4.

[25] KULLMAN, M. y CEFALIELLO, A.: "The interconnection between the AI Act and the EU's Occupational Safety and Health Legal Framework", January de 2022, disponible en http://global-workplace-law-and-policy.kluwerlawonline.com/2022/01/24/the-interconnection-between-the-ai-act-and-the-eus occupational-safety-and-health-legal-framework/

un riesgo significativo de daño a la salud, la seguridad o los derechos fundamentales de las personas físicas[26]. En todo caso, dichos proveedores estarán sujetos a la obligación de registro establecida en el artículo 49, apartado 2. A petición de las autoridades nacionales competentes, el proveedor facilitará la documentación de la evaluación (art. 6.4 RIA).

## 4. Implicaciones dañosas derivadas de la ia relacionadas con la seguridad y salud laboral

A la hora de valorar las implicaciones dañosas derivadas de los sistemas de IA, el intérprete debe tener en cuenta, asimismo, el distinto concepto de daño o perjuicio que lleva implícito la materialización del riesgo en cada una de las dos normativas de IA y de prevención de riesgos laborales.

En la normativa preventiva laboral, el concepto legal de daño derivado del riesgo laboral se concreta en "enfermedades, patologías o lesiones sufridas con motivo u ocasión del trabajo (art. 4.3 LPRL), englobando, también, los daños psicosociales que traen causa de la interacción del trabajador con la máquina.

Sin embargo, el concepto de daño en el RIA es más heterogéneo y plantea alguna otra dimensión añadida a la de seguridad y salud, relacionada con la posible vulneración de derechos fundamentales.

De entrada, toma en consideración los daños y perjuicios que sufran las personas o bienes como consecuencia del uso de los sistemas de IA. Los posibles perjuicios pueden ser resultado de defectos en el diseño general de los sistemas de IA, de un funcionamiento incorrecto, o de incidentes graves asociados al uso de sus sistemas de IA, y se concretan, bien en daños graves para la salud de las personas trabajadoras, o bien en una alteración grave e irreversible de la gestión o el funcionamiento de infraestructuras críticas, o daños graves a la propiedad o al medio ambiente.

Dentro de esta amplia gama de daños, se deben considerar los impactos emocionales, como la ansiedad y el estrés, la pérdida de control, el asilamiento si interactúan solo con sistemas de IA o la pérdida de significado o de propósito, que son los riesgos más habitualmente notificados con relación a la utilización de los sistemas de IA en el lugar de trabajo. El uso intensivo de sistemas de gestión de personas basadas en la IA que obliga a los trabajadores a trabajar más rápido o que les mantiene estar continuamente conectados, o contantemente vigilados o controlados, puede provocar elevados niveles de estrés laboral, ansiedad y depresión con los consiguientes efectos sobre la salud[27].

Pero el perjuicio considerado por el RIA va más allá, relacionándose, además, con el incumplimiento de obligaciones derivadas del Derecho de la Unión destinadas a pro-

[26] FERNÁNDEZ HERNÁNDEZ, C. Y EGUILUZ CASTAÑEIRA, J. A.: "Diez puntos críticos del Reglamento europeo de Inteligencia Artificial", *Diario La Ley*, op cit.

[27] EU-OSHA: "Inteligencia Artificial para la gestión de las personas trabajadoras: riesgos y oportunidades". 10/08/2022, Disponible en: https://osha.europa.eu/es/publications/artificial-intelligence-worker-management-risks-and-opportunities.

teger los derechos fundamentales, señaladamente con el tratamiento de los datos digitales de las personas. El uso de los sistemas de IA puede deparar consecuencias perjudiciales de control invasivo, de discriminación o de uso ilícito de datos personales, de forma que ciertas personas pueden ver vulnerada su intimidad, o ser discriminadas en las valoraciones, ascensos, o extinciones, por hacer suposiciones en función de sus características.

En este sentido, incorpora una dimensión que podríamos calificar de "carácter moral"[28], puesto que se valoran también las consecuencias adversas de un sistema de IA para los derechos fundamentales de las personas de la Unión, protegidos por la Carta de Derechos de la UE. Se mencionan, entre otros, 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, y en especial la no discriminación (Considerando 48), que no interesan desde un punto de vista estrictamente preventivo de seguridad y salud laboral.

Puede suceder que determinados algoritmos de la IA capturen patrones ocultos que reflejen prejuicios humanos como el racismo, el sexismo, la discriminación por edad. Los sistemas de IA empleados, por ejemplo, para controlar el rendimiento y el comportamiento de las personas pueden acabar socavando los derechos fundamentales a la protección de datos personales y a la intimidad" (Considerando 57), causando un perjuicio moral por vulneración de derechos fundamentales y no físico, que dé lugar una indemnización de daños y perjuicios.

## 5. Obligaciones preventivas laborales del proveedor de sistemas de IA

Ya se ha comentado que el RIA ha tratado de conectarse con el Derecho de la Unión en materia relativa a las condiciones de empleo y de trabajo, incluidas la salud y seguridad en el trabajo (Considerando 9)[29], por lo que debe ser interpretado en consonancia con la normativa de prevención de riesgos laborales, en particular, Directiva 89/391 CEE, Marco de seguridad y salud laboral y la normativa interna de trasposición (LPRL).

A fin de proteger los derechos de los trabajadores frente a los riesgos de seguridad y salud derivados del uso de los sistemas de IA en el lugar de trabajo, los proveedores deben ser capaces de incorporar a sus diseños y desarrollos los aspectos de prevención de riesgos laborales como cuestión de interés público, al igual que otras medidas en lo que sea estrictamente necesario para garantizar la detección y corrección de los sesgos asociados a los sistemas de IA de alto riesgo, con sujeción a las garantías adecuadas para los derechos y libertades fundamentales de las personas.

[28] RODRÍGUEZ SANZ DE GALDEANO, B.: "La responsabilidad empresarial por accidentes vinculados a la Inteligencia Artificial", *Trabajo y Derecho*, nº 19, junio 2024.

[29] Considerando 9: "*Además, en el contexto del empleo y la protección de los trabajadores, el presente Reglamento no debe afectar, por tanto, al Derecho de la Unión en materia de política social ni a la legislación laboral nacional —conforme al Derecho de la Unión— relativa a las condiciones de empleo y de trabajo, incluidas la salud y seguridad en el trabajo y la relación entre empleadores y trabajadores*".

No obstante, las obligaciones del proveedor de sistemas de IA varían según la clase de IA y el tipo de riesgo que comportan. El RIA establece, por un lado, distintas normas específicas para los modelos de IA de uso general sin riesgos sistémicos y para los modelos de IA con riesgos sistémicos, que deben aplicarse también cuando estos modelos estén integrados en un sistema de IA o formen parte de un sistema de IA; y, por otro lado, normas específicas para los sistemas de IA.

### *5.1. Obligaciones de los proveedores de modelos de IA de uso general con riesgo sistémico*

Es preciso diferenciar el concepto de modelos de IA de uso general del concepto de sistemas de IA con el fin de garantizar la seguridad jurídica. Los modelos de IA son componentes esenciales de los sistemas de IA, no constituyen por sí mismos sistemas de IA. Los modelos de IA requieren que se les añadan otros componentes, como, por ejemplo, una interfaz de usuario, para convertirse en sistemas de IA. Los modelos de IA suelen estar integrados en los sistemas de IA y formar parte de dichos sistemas (Considerando 97).

Por otra parte, dentro de los modelos de IA de uso general debe diferenciarse entre los modelos de IA sin riesgo sistémico y modelos de IA con riesgo sistémico, que conllevan, por ejemplo, "cualquier efecto negativo real o razonablemente previsible en relación con accidentes graves, perturbaciones de sectores críticos y consecuencias graves para la salud y la seguridad públicas, cualquier efecto negativo real o razonablemente previsible sobre los procesos democráticos y la seguridad pública y económica o la difusión de contenidos ilícitos, falsos o discriminatorios" (Considerando 110).

Los modelos de IA de uso general sin riesgo sistémico deben cumplir una serie de obligaciones y requisitos establecidos en el artículo 53 RIA. Son fundamentalmente medidas de transparencia proporcionadas, lo que incluye elaborar documentación y mantenerla actualizada y facilitar información sobre el modelo de IA de uso general, incluida la información relativa al proceso de entrenamiento y realización de pruebas y los resultados de su evaluación, para su uso por parte de los proveedores posteriores.

El proveedor del modelo de IA de uso general tiene la obligación de elaborar y mantener actualizada la documentación técnica con el fin de ponerla a disposición, previa solicitud, de la Oficina de IA y de las autoridades nacionales competentes. Esta información debe permitir a los proveedores de sistemas de IA entender bien las capacidades y limitaciones del modelo de IA de uso general y cumplir sus obligaciones. Los elementos mínimos que debe contener dicha documentación se encuentran en los anexos XI y XII específicos del Reglamento.

Un modelo de IA de uso general presenta riesgos sistémicos cuando tiene capacidades de gran impacto —evaluadas mediante herramientas y metodologías técnicas adecuadas— o unas repercusiones considerables en el mercado interior debido a su alcance. La cantidad acumulada de cálculo utilizado para el entrenamiento del modelo de IA de uso general, medida en operaciones de coma flotante, es una de las aproximaciones pertinentes para las capacidades del modelo. Cuando se alcanza un umbral inicial de ope-

raciones de coma flotante se presume que el modelo es un modelo de IA de uso general con riesgos sistémicos (Considerando 111).

La Comisión puede adoptar decisiones individuales por las que se designe un modelo de IA de uso general como modelo de IA de uso general con riesgo sistémico, atendiendo a una evaluación global de los criterios para la designación de modelos de IA de uso general con riesgo sistémico establecidos en un anexo del presente Reglamento, como la calidad o el tamaño del conjunto de datos de entrenamiento, el número de usuarios profesionales y finales, sus modalidades de entrada y de salida, su nivel de autonomía y escalabilidad o las herramientas a las que tiene acceso. Y ello particularmente cuando descubre que un modelo de IA de uso general del que no tenía conocimiento o que el proveedor pertinente no le había notificado cumple los requisitos para ser clasificado como modelo de IA de uso general con riesgo sistémico (Considerando 113)

Los modelos de IA de uso general con riesgo sistémico están sujetos, además de a las obligaciones impuestas a los proveedores de modelos de IA de uso general, a las específicas establecidas en el artículo 55 del Reglamento. Estas incluyen obligaciones encaminadas a detectar y atenuar dichos riesgos y a garantizar un nivel adecuado de protección en materia de ciberseguridad, independientemente de si dichos modelos se ofrecen como modelos independientes o están integrados en sistemas de IA o en productos

Además, los proveedores de estos modelos de IA deben realizar una evaluación de riesgos y mitigar continuamente los riesgos sistémicos, por ejemplo, mediante el establecimiento de políticas de gestión de riesgos, como procesos de rendición de cuentas y gobernanza, la puesta en práctica de la vigilancia poscomercialización, la adopción de medidas adecuadas durante todo el ciclo de vida del modelo y la cooperación con los agentes pertinentes a lo largo de la cadena de valor de la IA. Si, a pesar de los esfuerzos por detectar y prevenir los riesgos, el desarrollo o el uso del modelo provoca un incidente grave, el proveedor del modelo de IA de uso general debe, sin demora indebida, hacer un seguimiento del incidente y comunicar toda la información pertinente y las posibles medidas correctoras a la Comisión y a las autoridades nacionales competentes (Considerando 115).

Por lo que respecta a las obligaciones preventivas de seguridad y salud laboral, hay que tener en cuenta que cuando los modelos de IA se integran en sistemas de IA se deben seguir aplicando, además de las obligaciones establecidas en el Reglamento IA en relación con los modelos de IA, las establecidas en relación con los sistemas de IA. Por tanto respecto de las obligaciones en materia preventiva hay que estar a lo que a continuación se indicará sobre las obligaciones preventivas del proveedor de los sistemas de IA de alto riesgo.

### *5.2. Obligaciones de proveedores de sistemas de IA de alto riesgo*

Con carácter general, los proveedores que comercialicen o pongan en servicio sistemas de IA de alto riesgo en la Unión, con independencia de que dichos proveedores estén

establecidos o ubicados en la Unión o un tercer país, se encuentran obligados a observar una serie de requisitos previstos en el art. 16 del capítulo III.

Entre ellos, el primero y básico es que los sistemas de IA de alto riesgo cumplan los requisitos establecidos sección 2 del capítulo III , "*teniendo en cuenta sus finalidades previstas y el estado actual de la técnica generalmente reconocido en materia de IA*" (art. 8 RIA). Los requisitos de los sistemas de IA de alto riesgo establecidos en la sección 2 del capítulo III, se resumen básicamente en los siguientes: establecimiento, documentación y mantenimiento de un sistema de gestión de riesgos (art. 9) (incluida la evaluación previa de conformidad); aseguramiento de la alta calidad de datos utilizados en el entrenamiento del sistema para minimizar riesgos y resultados discriminatorios (art. 10); disponer de una precisa documentación que cumpla con los requisitos del Anexo IV (art. 11); garantizar la trazabilidad del funcionamiento del sistema mediante un registro de actividad (art. 12); ofrecer una información clara y adecuada al responsable del despliegue para que lo interprete y use correctamente (art. 13)[30]; contar con medidas de vigilancia humana, dotándolos de herramientas de interfaz humano-máquina para prevenir o reducir al mínimo los riesgos que pueden surgir del uso adecuado a la finalidad

[30] Art. 13 RIA*: "Las instrucciones de uso contendrán al menos la siguiente información:*

*a) la identidad y los datos de contacto del proveedor y, en su caso, de su representante autorizado;*

*b) las características, capacidades y limitaciones del funcionamiento del sistema de IA de alto riesgo, y en particular:*

*i) su finalidad prevista;*

*ii) el nivel de precisión (incluidos los parámetros para evaluarla), solidez y ciberseguridad mencionado en el artículo 15 con respecto al cual se haya probado y validado el sistema de IA de alto riesgo y que puede esperarse, así como cualquier circunstancia conocida y previsible que pueda afectar al nivel de precisión, solidez y ciberseguridad esperado;*

*iii) cualquier circunstancia conocida o previsible, asociada a la utilización del sistema de IA de alto riesgo conforme a su finalidad prevista o a un uso indebido razonablemente previsible, que pueda dar lugar a riesgos para la salud y la seguridad o los derechos fundamentales a que se refiere el artículo 9, apartado 2;*

*iv) en su caso, las capacidades y características técnicas del sistema de IA de alto riesgo para proporcionar información pertinente para explicar su información de salida;*

*v) cuando proceda, su funcionamiento con respecto a personas o grupos de personas específicos en relación con los que esté previsto utilizar el sistema;*

*vi) cuando proceda, especificaciones relativas a los datos de entrada, o cualquier otra información pertinente en relación con los conjuntos de datos de entrenamiento, validación y prueba usados, teniendo en cuenta la finalidad prevista del sistema de IA;*

*vii) en su caso, información que permita a los responsables del despliegue interpretar la información de salida del sistema de IA de alto riesgo y utilizarla adecuadamente;*

*c) los cambios en el sistema de IA de alto riesgo y su funcionamiento predeterminados por el proveedor en el momento de efectuar la evaluación de la conformidad inicial, en su caso;*

*d) las medidas de vigilancia humana a que se hace referencia en el artículo 14, incluidas las medidas técnicas establecidas para facilitar la interpretación de la información de salida de los sistemas de IA de alto riesgo por parte de los responsables del despliegue;*

*e) los recursos informáticos y de hardware necesarios, la vida útil prevista del sistema de IA de alto riesgo y las medidas de mantenimiento y cuidado necesarias (incluida su frecuencia) para garantizar el correcto funcionamiento de dicho sistema, también en lo que respecta a las actualizaciones del software;*

*f) cuando proceda, una descripción de los mecanismos incluidos en el sistema de IA de alto riesgo que permitir a los responsables del despliegue recabar, almacenar e interpretar correctamente los archivos de registro de conformidad con el artículo 12*".

como del uso indebido razonablemente previsible (art. 14); ofrecer un nivel adecuado de precisión, solidez, y ciberseguridad durante todo su ciclo de vida, adoptando las medidas técnica y organizativas necesarias (art. 15).

A ello se unen las obligaciones de: colocar en el embalaje del sistema de IA su nombre comercial y dirección de contacto; disponer de un sistema de gestión de calidad (art. 17); conservar la documentación relativa a técnica, sistema de gestión de la calidad, cambios aprobados, declaración UE de conformidad ( art. 18); conservar los archivos de registro generados automáticamente por sus sistemas de IA de alto riesgo a que se refiere el artículo 19; asegurar que los sistemas de IA de alto riesgo sean sometidos al procedimiento pertinente de evaluación de la conformidad antes de su introducción en el mercado o puesta en servicio (ex art 43); elaborar una declaración UE de conformidad (ex art. 47); colocar en el sistema de IA de alto riesgo, en su embalaje o documentación el marcado CE de conformidad con el presente Reglamento (ex art. 48); cumplir las obligaciones de registro (ex art. 49.1); establecer medidas correctoras cuando tenga motivos para ello, retirando desactivando o recuperando; velar por que el sistema de IA de alto riesgo cumpla requisitos de accesibilidad productos y servicios.

Lo señalado solo es de aplicación, como ya se ha indicado antes, a los proveedores que despliegan un sistema de IA que merezca la consideración de alto riesgo por entrañar un riesgo considerable de causar un perjuicio a los intereses jurídicos amparados por el RIA. En consecuencia, de no influir sustancialmente en la toma de decisiones ni perjudicar dichos intereses sustancialmente", no resulta de aplicación. Así lo remarca, por otra parte, el art. 6.3 RIA: "N*o obstante lo dispuesto en el apartado 2, un sistema de IA no se considerará de alto riesgo si no plantea un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas, en particular al no influir sustancialmente en el resultado de la toma de decisiones.*

En particular, el artículo 6.3 RIA considera que los sistemas de IA no entrañan dicho riesgo importante, y, en consecuencia, no constituyen alto riesgo cuando *se cumplan una o varias de las condiciones siguientes:*

*a) que el sistema de IA tenga por objeto llevar a cabo una tarea de procedimiento limitada;*
*b) que el sistema de IA tenga por objeto mejorar el resultado de una actividad humana previamente realizada;*
*c) que el sistema de IA tenga por objeto detectar patrones de toma de decisiones o desviaciones con respecto a patrones de toma de decisiones anteriores y no esté destinado a sustituir la evaluación humana previamente realizada sin una revisión humana adecuada, ni a influir en ella; o*
*d) que el sistema de IA tenga por objeto llevar a cabo una tarea preparatoria para una evaluación pertinente a efectos de los casos de uso enumerados en el anexo III".*

De todas formas, es esencial indicar que un sistema de IA será considerado de alto riesgo si realiza una elaboración de perfiles o perfilado de personas física, dejando

inoperativas las excepciones anteriormente mencionadas[31]. Dispone, en este sentido, el art. 6.3 RIA, en su último párrafo, que: "*No obstante lo dispuesto en el párrafo primero, los sistemas de IA a que se refiere el anexo III siempre se considerarán de alto riesgo cuando el sistema de IA lleve a cabo la elaboración de perfiles de personas físicas*". Este elemento introduce "una dinámica potencialmente conflictiva donde los sistemas de IA, de otro modo evaluados como de bajo riesgo, pueden categorizarse en un estatus de alto riesgo únicamente debido a las características integradas, independientemente de su impacto real de riesgo real"[32].

Nos encontramos, así, con una sucesión encadenada de reglas técnicas, bastante abstractas, de excepción y contra excepción, que atenúan y debilitan la percepción de los principios y garantías que deben cumplir los sistemas de IA de alto riesgo. Su inespecificidad hace complicado y confuso anticipar cuándo se está ante un sistema de IA sujeto a las obligaciones aplicables con carácter general a los sistemas de IA de alto riesgo.

No obstante, de lo expuesto se deriva que puede haber sistemas de IA de alto riesgo que, aun catalogados como tales, no lo son porque no entrañan riesgos importantes, y otros que no siendo de alto riesgo por no causar daño considerable, pueden ser, sin embargo, catalogados como tales, porque así lo ha determinado el RIA, que serían los casos en los que se lleve a cabo la elaboración de perfiles de personas físicas.

En todo caso, el proveedor siempre podrá disipar las posibles dudas de aplicación de la serie de obligaciones que contempla la sección 2 del Capítulo III, acogiéndose a alguno de los mecanismos de acreditación de conformidad de los sistemas de IA de alto riesgo, y con ello podrá conjurar cualquier eventual responsabilidad por no haber adoptado dichas medidas.

El proveedor puede optar por aplicar las normas armonizadas publicadas en el DOUE a que se refiere el artículo 40, o bien, en su caso, las especificaciones comunes emitidas por organismos a que se refiere el artículo 41, o el procedimiento de evaluación de la conformidad, basado en el procedimiento interno de control, establecido en el anexo VI, o los certificados emitidos por los organismos competentes (ex art. 44) que serán válidos para el período que indiquen, que no excederá de cuatro años para los sistemas de IA contemplados en el anexo III[33].

Por tanto, una posibilidad, entre otras, sería diseñar los sistemas de IA siguiendo las normas armonizadas del art. 40 , y someterse a los correspondientes procedimientos de evaluación de conformidad (art. 43). El art. 40 establece que "*se presumirá que los sistemas de IA de alto riesgo que sean conformes con normas armonizadas, o partes de estas, cuyas referencias estén publicadas en el Diario Oficial de la Unión Europea de conformidad con el Reglamento (UE) n.º 1025/2012 son conformes con los requisitos establecidos en la*

[31] FERNÁNDEZ HERNÁNDEZ, C. Y EGUILUZ CASTAÑEIRA, J. A.: "Diez puntos críticos del Reglamento europeo de Inteligencia Artificial", *Diario La Ley,* op cit.

[32] Ibidem.

[33] A solicitud del proveedor, la validez de un certificado podrá prorrogarse por períodos adicionales no superiores a cuatro años para los sistemas de IA contemplados en el anexo III, sobre la base de una nueva evaluación con arreglo a los procedimientos de evaluación de la conformidad aplicables.

*sección 2 del presente capítulo o, en su caso, con las obligaciones establecidas en el capítulo IV del presente Reglamento, en la medida en que dichas normas abarquen estos requisitos u obligaciones*". El problema radica en que no se cuenta todavía con esas normas técnicas actualizadas que recojan los estándares técnicos de seguridad[34].

Adentrándonos en el terreno específico de las obligaciones preventivas laborales del proveedor, es preciso distinguir entre los sistemas de IA sujetos a los requisitos de alto riesgo y los que no están obligados a ello.

1) En los sistemas de IA de alto riesgo, destinados a ser utilizados en el trabajo, que cumplan los requisitos establecidos sección 2 del capítulo III y entrañen un riesgo considerable de causar un perjuicio a los intereses jurídicos amparados por el RIA, y en aquellos otros sistemas de IA que, sin comportar un riesgo importante, "*se consideran de alto riesgo por llevar a cabo la elaboración de perfiles de personas físicas"*, la previsión de riesgos laborales constituye una parte esencial de la declaración de conformidad, porque según lo previsto en el art. 8 RIA, deben tener en cuenta "*sus finalidades previstas y el estado actual de la técnica generalmente reconocido en materia de IA*".

Tanto si se trata de sistema de IA de alto riesgo destinado a ser utilizado como componente de una máquina, como si se considera un producto en sí mismo, siempre que sean diseñados para ser utilizados en el trabajo, el proveedor está obligado a garantizar que el sistema de IA de alto riesgo ha sido entrenado, validado y probado con datos que reflejen el entorno geográfico, conductual contextual o funcional específico en el que esté previsto su uso (art. 42 RIA).

Dichas prácticas habrán de realizarse teniendo en cuenta, según el art. 10 RIA, el diseño, los procesos de recogida de datos, la finalidad original de la recogida, las operaciones de tratamiento de datos, la formulación de supuestos, en particular en lo que respecta a la información que se supone que miden y representan los datos; 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 algún tipo de discriminación prohibida por el Derecho de la Unión, especialmente cuando las salidas de datos influyan en las informaciones de entrada de futuras operaciones; y las medidas adecuadas para detectar, prevenir y reducir posibles sesgos detectados.

Es decir, el proveedor deberá tener muy presente la finalidad del sistema de IA y los datos que se van a recabar, deberá entrenar con datos que reflejen el entorno en que se va a utilizar el sistema de IA, y prever los posibles riesgos para la seguridad y salud laboral y la violación de los derechos fundamentales, así como estimar la magnitud de los mismos, y tomar una decisión apropiada sobre las medidas que deben adoptarse de prevención de riesgos laborales.

[34] RODRÍGUEZ SANZ DE GALDEANO, B.: "La responsabilidad empresarial por accidentes vinculados a la Inteligencia Artificial", *Trabajo y Derecho*, nº 19, junio 2024.

El RIA ha tratado de conectar los requisitos del Reglamento IA con otras normas sectoriales de la Unión, indicando, como ya se ha reiterado, que "*completan el conjunto existente de actos de armonización de la Unión*". En este sentido, los sistemas de IA destinados a ser integrados en máquinas, habrán de ajustarse a las especificaciones que indica la normativa de máquinas, en concreto, el Reglamento (UE) 2023/1230, relativo a máquinas y por el que se deroga la Directiva 2006/42/CE y la Directiva 73/361/CEE, (no aplicable hasta el 20 de enero de 2027), que ha venido precisamente a regular los nuevos riesgos emergentes de la Inteligencia artificial y el Internet de las cosas (IoT). De igual modo, en el supuesto de sistemas de IA destinados a ser incorporados a equipos de protección individual, será necesario que cumplan con su normativa específica; en este caso, el Reglamento 2016/425 que prevé los requisitos esenciales de seguridad que han de reunir estos equipos. Y todo ello sin perjuicio del deber de ajustarse a la normativa de protección de datos, aspecto no abordado en este estudio [35]

A esta premisa acompaña luego, como ulterior corolario, el deber de todo fabricante de productos o equipos de cumplir, de acuerdo con lo dispuesto en el art. 41 LPRL, las obligaciones generales de seguridad. Aquí pueden surgir algunas fricciones, porque la primera obligación del fabricante es que la maquinaria, los equipos, productos y útiles de trabajo "no constituyan una fuente de peligro" (art. 41.1 LPRL), y el RIA no garantiza, sin embargo, el mismo nivel de protección por cuanto no excluye un riesgo alto si no causa un perjuicio considerable, como hemos señalado. Ocurre, además, que los valores del RIA no son solo de seguridad sino también de mercado; y por tanto la comercialización se ha convertido en rasgo esencial. Con lo cual, la aplicación íntegra de las obligaciones del art. 41 LPRL al proveedor de sistemas de IA resulta insostenible y obliga a una afinación o reformulación de las mismas, reconociendo su relatividad respecto al modo de concretarse en el trabajo.

2) En el caso de los sistemas de IA que no representan un alto riesgo (riesgo medio o bajo), los requisitos de seguridad previstos en el RIA se traducen generalmente en obligaciones de transparencia o información que permitan a los usuarios ser conscientes de que interactúan con un sistema de IA y comprender sus características y limitaciones. Así se establece, por ejemplo, como ya se apuntado anteriormente, con respecto a los sistemas de IA diseñados para interactuar con personas (chatbots); o en el caso de los sistemas de IA, incluidos los sistemas de uso general, debiendo garantizar que los resultados del sistema

[35] Considerando 70: *A fin de proteger los derechos de terceros frente a la discriminación que podría provocar el sesgo de los sistemas de IA, los proveedores deben —con carácter excepcional, en la medida en que sea estrictamente necesario para garantizar la detección y corrección de los sesgos asociados a los sistemas de IA de alto riesgo, con sujeción a las garantías adecuadas para los derechos y libertades fundamentales de las personas físicas y tras la aplicación de todas las condiciones aplicables establecidas en el presente Reglamento, además de las condiciones establecidas en los Reglamentos (UE) 2016/679 y (UE) 2018/1725 y la Directiva (UE) 2016/680— ser capaces de tratar también categorías especiales de datos personales, como cuestión de interés público esencial en el sentido del artículo 9, apartado 2, letra g), del Reglamento (UE) 2016/679 y del artículo 10, apartado 2, letra g), del Reglamento (UE) 2018/1725*.

> de IA estén marcados en un formato legible por máquina y detectable como generado o manipulado artificialmente; o con relación a los sistemas de reconocimiento de emociones y sistemas de clasificación biométrica, obligando a informar de que se están usando; o en los sistemas que generen o manipulen imágenes o audio de personas (ultrafalsificación) teniendo que informar de que se trata de imágenes o sonidos manipulados artificialmente.

Es poco probable el empleo de estos sistemas de IA en el trabajo, pero, de ser utilizados en el trabajo, en principio, les resultaría de aplicación las obligaciones preventivas laborales previstas genéricamente en el art. 41 de la LPRL para el fabricante de productos de trabajo. Como en el caso anterior, el proveedor de tales sistemas vendría obligado a: asegurar que éstos no constituyen una fuente de peligro para el trabajador; a envasar y etiquetar correctamente de forma que se permita su conservación y manipulación en condiciones de seguridad, e identificar claramente los riesgos para la seguridad y salud de los trabajadores que su utilización comporten; y a proporcionar información sobre la forma correcta de utilización las medidas preventivas adicionales y los riesgos laborales que conlleva tanto su uso normal como su manipulación o empleo inadecuado[36].

Ahora bien, en tales casos volverían a chocar los fines preventivos laborales con los del RIA que aspira también a facilitar la libre circulación y el reconocimiento de los sistemas de IA y a promover el comercio internacional de la industria de la IA. No está, por tanto, muy claro que los proveedores de estos sistemas de IA de riesgo medio o bajo estén obligados a cumplir estrictamente cada una de las exigencias del art. 41 LPRL, dado que ello daría lugar, además, al establecimiento de unas garantías de exigencia de seguridad más altas, que en los de riesgo alto, lo que carece de toda lógica.

## 6. Obligaciones preventivas del empresario respecto de los sistemas de IA

El Reglamento de IA prevé también obligaciones para los responsables de la implantación de los sistemas de IA de alto riesgo, esto es, para los empresarios en cuanto usuarios o implementadores de tales productos que tengan su lugar de establecimiento, o estén situados, en la Unión (art. 2.1 RIA). Ellos conocen mejor que nadie el uso concreto que se le dará al sistema de IA de alto riesgo y pueden, por lo tanto, *"detectar potenciales riesgos significativos que no se previeron en la fase de desarrollo, al tener un conocimiento más preciso del contexto de uso y de las personas o los grupos de personas que probablemente se vean afectados, entre los que se incluyen grupos de personas vulnerables*" (Considerando 93).

Las obligaciones impuestas al empresario por el art. 26 del RIA tienen un carácter marcadamente instrumental en la medida que aspiran a garantizar fundamentalmente la correcta aplicación del sistema de IA de alto riesgo, conforme a las indicaciones de uso dadas por el proveedor o fabricante del mismo. Dentro del amplio elenco de obli-

[36] FRANCIS LEFEBVRE: *Memento Social. Prevención de Riesgos Laborales*, *2024-2025*, Madrid , p. 1416.

gaciones, es posible distinguir entre: A) las obligaciones previas a la puesta en servicio o utilización del sistema de IA del alto riesgo (obligaciones de diseño), y B) las relativas al despliegue o uso del mismo (obligaciones de uso).

A) Obligaciones de diseño. Antes de poner en servicio, el empresario se encuentra obligado a: 1) adoptar las medidas técnicas y organizativas adecuadas para garantizar que utilizan dichos sistemas de conformidad con sus instrucciones de uso; 2) encomendar la supervisión humana a personas físicas con competencia y formación necesarias; 3) asegurarse de que los datos de entrada sean pertinentes y suficientemente representativos para la finalidad prevista del sistema de IA de alto riesgo, en la medida en que ejerza el control sobre dichos datos, 4) e informar 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 (art. 26. 1, 2, 4 y 7 RIA). Pero para ello el empresario debe estar correctamente informado de las características, las capacidades y las limitaciones del funcionamiento del sistema de IA.

A tal fin, el Reglamento exige al fabricante transparencia respecto de los sistemas de IA de alto riesgo, de modo que permita al empresario comprender la manera en que el sistema de IA funciona, evaluar su funcionalidad y comprender sus fortalezas y limitaciones. Estos elementos abarcarían, según precisa el Considerando 72, *la información sobre las posibles circunstancias conocidas o previsibles relacionadas con el uso del sistema de IA de alto riesgo, incluida la actuación del responsable del despliegue capaz de influir en el comportamiento y el funcionamiento del sistema, en cuyo marco el sistema de IA puede dar lugar a riesgos para la salud, la seguridad y los derechos fundamentales, sobre los cambios que el proveedor haya predeterminado y evaluado para comprobar su conformidad y sobre las medidas pertinentes de supervisión humana, incluidas las medidas para facilitar la interpretación de la información de salida del sistema de IA por parte de los responsables del despliegue. La transparencia, incluidas las instrucciones de uso que acompañan a los sistemas de IA, debe ayudar a los responsables del despliegue a utilizar el sistema y tomar decisiones con conocimiento de causa*.

Del conjunto de obligaciones señaladas, se desprende que el empresario asume en relación con la adopción de la IA de alto riesgo los siguientes compromisos (de los que cabe colegir eventualmente responsabilidades): el primero, estar informado sobre los usos previstos y excluidos; el segundo, elegir correctamente el sistema de IA de alto riesgo en función de los usos previstos y no excluidos (la adopción de estos sistemas de IA debe responder a un fin permitido)[37]; el tercero, garantizar que el personal que se encargue de la supervisión tiene el nivel de formación necesario (la alfabetización en materia de inteligencia artificial y los conocimientos necesarios para garantizar el cumplimiento adecuado

[37] Considerando 74 RIA : *El nivel previsto de los parámetros de funcionamiento debe declararse en las instrucciones de uso que acompañen a los sistemas de IA*.

y la correcta ejecución la debe proporcionar el empresario)[38]; y el cuarto, informar tanto a la representación legal de los trabajadores como a los trabajadores afectados de los conocimientos necesarios para garantizar el cumplimiento adecuado y la correcta ejecución.

Sobre estas obligaciones se interfieren las derivadas de la normativa de protección de datos (RGPD y LOPDGDD), que, entre otras exigencias, impone realizar una evaluación de impacto sobre los derechos fundamentales de los sistemas de IA de alto riesgo (art. 27 RIA) (asunto que merece un tratamiento específico), pero cuyos resultados bien pudieran limitar la adopción de los sistemas de IA si esta tecnología entra en conflicto con dicha normativa[39]. El cumplimiento de las obligaciones del RIA aplicables a los sistemas de IA debe entenderse sin perjuicio de otras obligaciones a los responsables del despliegue de sistemas de IA establecidas en el Derecho de la Unión o nacional, como las señaladas de protección de datos personales.

B) Obligaciones de uso. Una vez implantado el sistema de IA de alto riesgo, el empleador debe ejercer un control sobre los datos de entrada, garantizar que dichos datos sean pertinentes y suficientemente representativos a la vista de la finalidad prevista del sistema, supervisar el funcionamiento del sistema sobre la base de las instrucciones de uso , informar al proveedor cuando el sistema presente un riesgo, suspendiendo el uso del sistema, y conservar los registros generados automáticamente por el sistema durante un periodo adecuado a la finalidad prevista del sistema de IA de alto riesgo, de al menos seis meses (art. 26. 3,4, 5 y 6 RIA)

A ello se añade la obligación del empleador que usa el sistema de IA de alto riesgo de los incluidos en el Anexo III, de informar a las personas físicas (trabajadores) de que están expuestas a la utilización de los sistemas de la IA de alto riesgo (art. 26.7 RIA). Y si, además, se ven afectadas por una decisión que el responsable adopte basándose en los resultados de un sistema de IA de alto riesgo y que produzca efectos jurídicos o le afecte considerablemente del mismo modo, de forma que considere que tiene un efecto perjudicial para la salud, su seguridad o sus derechos fundamentales, tendrán derecho a obtener del empresario explicaciones claras y significativas acerca del papel que el sistema ha tenido en el proceso de toma de decisiones y los principales elementos de la decisión adoptada.

Anótese además, que los responsables del despliegue que utilicen un sistema de IA para generar o manipular un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeje notablemente a personas, lugares o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos (ultrafalsificaciones) están

[38] Considerando 91 RIA: *Los responsables del despliegue deben garantizar que las personas encargadas de poner en práctica las instrucciones de uso y la supervisión humana establecidas en el presente Reglamento tengan las competencias necesarias, en particular un nivel adecuado de alfabetización, formación y autoridad en materia de IA para desempeñar adecuadamente dichas tareas*.

[39] RODRÍGUEZ SANZ DE GALDEANO, B.: La responsabilidad empresarial por accidentes vinculados a la Inteligencia Artificial", Trabajo y Derecho, op. cit.

obligadas también a hacer público, de manera clara y distinguible, que este contenido ha sido creado o manipulado, de manera artificial etiquetando la información de salida generada por la inteligencia artificial en consecuencia e indicando su origen artificial. En tales casos, el deber de transparencia en relación con las ultrafalsificaciones obliga al empresario a revelar la existencia de tales contenidos generados o manipulados (Considerando 134).

El deber empresarial de garantizar, conforme al RIA, la seguridad y salud en los sistemas de IA de alto riesgo es correlativo al deber de protección de los proveedores. Por tanto, sus obligaciones específicas como responsable de la implantación son exigibles únicamente respecto de aquellos sistemas de IA que cumplan los requisitos establecidos en la sección 2 del capítulo III y entrañen un riesgo considerable de causar un perjuicio a los intereses jurídicos amparados por el RIA y aquellos otros sistemas de IA que, sin comportar un riesgo importante, se consideran de alto riesgo por llevar a cabo la elaboración de perfiles de personas físicas.

No obstante, el alcance de sus obligaciones preventivas no se agota en la dimensión prevista por el RIA, sino que se integra al mismo tiempo por el conjunto de obligaciones propias de prevención de riesgos laborales, que el empresario, como garante de la seguridad y salud laboral de las personas trabajadoras, debe asumir, conforme a la LPRL. Ni las obligaciones de los proveedores ni las obligaciones propias suyas derivadas del RIA, eximen al empresario de cumplir con las obligaciones específicas en materia de prevención de riesgos laborales.

Desde esta otra perspectiva preventiva laboral, y de acuerdo con la LPRL, el empresario debe garantizar la seguridad y salud de los trabajadores a su servicio, adoptando cuantas medidas sean necesarias y desarrollando una acción permanente de seguimiento de la actividad preventiva. En concreto, el empresario habrá de identificar, evaluar y controlar los riesgos que no se hayan podido evitar, adoptar las medidas preventivas adecuadas en el lugar de trabajo, así como informar a los trabajadores sobre los riesgos que no se hayan podido evitar y formarles para el adecuado manejo de los equipos[40].

Y todo ello con respecto de cualquier dispositivo de IA y. no solo con relación a los sistemas de IA de alto riesgo que, según el RIA, entrañan un riesgo considerable de causar un perjuicio a los intereses jurídicos amparados por el RIA y aquellos otros sistemas de IA que, sin comportar un riesgo importante, se consideran de alto riesgo por llevar a cabo la elaboración de perfiles de personas física, dado el deber del empresario de adoptar cuantas medidas sean necesarias para evitar el daño cualquiera que sea.

## 7. Conclusiones

El Reglamento de IA tiene en común con la LPRL el elemento esencial del riesgo, pues ha sido diseñado bajo el enfoque del riesgo que la puesta en marcha y uso de determina-

[40] RODRÍGUEZ SANZ DE GALDEANO, B.: La responsabilidad empresarial por accidentes vinculados a la Inteligencia Artificial", Trabajo y Derecho, op. cit.

dos sistemas de IA plantean para la salud, la seguridad y derechos fundamentales consagrados en la Carta. Ambos comparten también el objeto porque, mientras la LPRL trata de identificar y estimar la magnitud del riesgo para posteriormente evitarlo o, en su caso, reducirlo y controlarlo[41], el RIA impone a los proveedores obligaciones orientadas a evaluar riesgos concretos y a aplicar medidas de reducción del riesgo razonable; su control constituye, por tanto, el objetivo clave de la actividad preventiva.

Pero hay elementos que separan a ambas normativas. Ante todo, la propia caracterización del riesgo. La LPRL acoge un criterio general, sin exclusiones o distinciones de categorías de riesgo; el deber de prevención laboral se extiende a cualquier tipo de riesgo. Sin embargo, el RIA limita sus exigencias a determinados tipos y categorías de riesgos. La Propuesta inicial de Reglamento de IA solo contemplaba la gama de "riesgos crónicos" (inaceptable, alto, limitado y mínimo), pero, a consecuencia de la aparición de los denominados GPAI, (los General Purpose AI sistems and models) se ha insertado una nueva regulación de los modelos y sistemas de IA de uso general, que puede plantear riesgos sistémicos. Por tanto, el RIA distingue entre dos grandes categorías: los riesgos sistémicos, de muy reducida probabilidad pero de una intensidad e implicaciones mucho más graves; y los riesgos crónicos de alta frecuencia pero de intensidad moderada.

Por otra parte, el riesgo tomado en consideración por el RIA se contrae a los sistemas de IA que presentan un "riesgo considerable" de ser perjudiciales para la salud y la seguridad o los derechos fundamentales de las personas, "teniendo en cuenta tanto la gravedad del posible perjuicio como la probabilidad de que se produzca". Junto a ello el RIA considera que la calificación de alto riesgo debe formularse previendo reducir al mínimo cualquier restricción del comercio internacional, pues tiene como objetivo promover el desarrollo responsable y sostenible de la IA en la Unión Europea, mediante la creación de un marco normativo armonizado y proporcional para la IA, que reduzca la fragmentación del mercado interno.

Todos estos condicionantes de magnitud del riesgo o de orden económico establecidos por el RIA sobre los sistemas de IA de alto riesgo hacen presagiar una limitada aplicación del RIA en el ámbito de las relaciones laborales por cuanto buena parte de los sistemas de IA utilizados en el ámbito laboral (la mayoría son de gestión de recursos humanos) no presentan de entrada un efecto nocivo grave o alto a los trabajadores, y porque el impacto negativo, muchas veces de carácter psicológico, no aparece de forma súbita sino de forma gradual. Ahora bien, no cabe obviar que los sistemas de IA, a pesar de no enmarcarse en el Anexo III también pueden categorizarse en un estatus de alto riesgo independientemente de su riesgo real, si realizan una elaboración de perfiles ex art. 6.3 RIA, por lo que es muy probable que muchos de los sistemas de IA utilizados actualmente en gestión de recursos humanos acaben siendo considerados sistemas de alto riesgo por esta vía de excepción.

[41] Así, entre los principios generales del deber general de prevención laboral (art. 15.1 LPRL), se hallan los de evitar los riesgos, evaluar los riesgos que no se puedan evitar, y combatir los riesgos en su origen

De todas formas, y con independencia de ello, el proveedor, que asume la mayor parte de las obligaciones preventivas que establece el RIA, debe tener en cuenta en el diseño y desarrollo de los sistemas de IA los aspectos de prevención de riesgos laborales como cuestión de interés público, porque el alcance de sus obligaciones preventivas no se agota en la dimensión prevista por el RIA, sino que se integra al mismo tiempo por el conjunto de obligaciones propias de prevención de riesgos laborales, que el fabricante, como garante de la seguridad y salud laboral de las personas trabajadoras, debe asumir, conforme al art. 41 de la LPRL.

Por otra parte, el empleador/usuario de los sistemas de IA en el lugar de trabajo debe tener presente que ni las obligaciones de los proveedores, ni las obligaciones propias suyas derivadas del RIA, le eximen de cumplir con las obligaciones específicas en materia de prevención de riesgos laborales que le impone la LPRL. Su deber empresarial de protección de la seguridad y salud laboral de las personas trabajadoras se superpone a las obligaciones del RIA, proyectándose, además, no solo sobre los sistemas de IA de alto riesgo que, según el RIA, entrañan un riesgo considerable de causar un perjuicio importante, sino también sobre cualquier dispositivo de IA con independencia de la categoría de riesgo que comporte.

## 8. Bibliografía consultada

ABDULLAH MALIK: *Artificial Intelligence in Health and Safety*, 22 /2/ 2023. Disponible en: https://safetypedia-com.translate.goog/safety/artificial-intelligence-in-health-and safety/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc

BARRIO ANDRÉS, M. : "Algunos claroscuros en el Reglamento Europeo de Inteligencia Artificial", *Diario LA LEY, Nº 86, Sección Ciberderecho*, 30 de Julio de 2024.

FERNÁNDEZ HERNÁNDEZ, C. y EGUILUZ CASTAÑEIRA, J. A.: "Diez puntos críticos del Reglamento europeo de Inteligencia Artificial", *Diario LA LEY, Nº 85, Sección Ciberderecho*, 28 de junio de 2024.

FRANCIS LEFEBVRE: *Memento Social. Prevención de Riesgos Laborales*, 2024-2025, Madrid , 2024, p. 1416.

GOÑI SEIN, J. L.: "El Reglamento UE de Inteligencia Artificial y su interrelación con la normativa de seguridad y salud en el trabajo", AA. VV. (Dirs. EGUSQUIZA, M. A.; RODRÍGUEZ SANZ DE GALDEANO, B.): *Inteligencia artificial y prevención de riesgos laborales: obligaciones y responsabilidades*, Tirant lo Blanch, Valencia 2023,

GOÑI SEIN,J.L.; RODRÍGUEZ SANZ DE GALDEANO, B.; LLORENS ESPADA, J.; MARIN MALO, M.: "El impacto del nuevo marco normativo europeo de la inteligencia artificial en las relaciones laborales", AA VV (Dir. RICHARD GONZÁLEZ, M.), en prensa edit J B. Bosch.

KULLMAN, M. y CEFALIELLO, A.: "The interconnection between the AI Act and the EU's Occupational Safety and Health Legal Framework", January de

2022, Disponible en http://global-workplace-law-and-policy.kluwerlawonline.com/2022/01/24/the-interconnection-between-the-ai-act-and-the-eusoccupational-safety-and-health-legal-framework/

LLORENS ESPADA, J.: *Límites al uso de la Inteligencia Artificial en el ámbito de la salud*, La Ley, Madrid, 2023, pp. 151 y ss.

LLORENS ESPADA, J.: "Inteligencia artificial y salud laboral", AA. VV. (Dirs. EGUSQUIZA, M. A.; RODRÍGUEZ SANZ DE GALDEANO, B.): *Inteligencia artificial y prevención de riesgos laborales: obligaciones y responsabilidades*, Tirant lo Blanch, Valencia 2023.

MERCADER UGUINA, J.: "Los usos de alto riesgo en el ámbito laboral de la IA y la certificación", *El Foro de Labos,* 9/5/2024, Disponible en: https://www.elforodelabos.es/2024/05/los-usos-de-alto-riesgo-en-el-ambito-laboral-de-la-ia-y-la-autocertificacion/

RODRÍGUEZ SANZ DE GALDEANO, B.: "La responsabilidad empresarial por accidentes vinculados a la Inteligencia Artificial", *Trabajo y Derecho,* nº 19, junio 2024.

VESTRI, G.: "La Unión Europea estrena el Reglamento de Inteligencia Artificial (RIA). Control, supervisión y uso de una tecnología cada vez más presente en la vida de todos", *Diario LA LEY*, Nº 10550, 19 de Julio de 2024.

# Daños derivados de la IA en el trabajo. Modelo regulador y responsabilidad civil[*]

## Damages derived from AI at work. Regulatory model and civil responsibility

Beatriz Rodríguez Sanz de Galdeano
*Profesora Titular de Derecho del Trabajo y de la Seguridad Social*
*Universidad Pública de Navarra*

doi: 10.20318/labos.2024.9037

*Resumen:* El objetivo del presente artículo es realizar una aproximación al modelo regulador que ha inspirado la nueva normativa en materia de IA y sus relaciones con el marco normativo específico en materia de prevención de riesgos laborales. Se trata de profundizar en las obligaciones generales de los proveedores de sistemas de IA y en las específicas, en materia de prevención de riesgos laborales, de tales proveedores y del empresario que incorpora dichos sistemas al ámbito laboral. Asimismo, se realiza un análisis del régimen vigente en materia de responsabilidad por daños, con el propósito de plantear el abanico de posibilidades de reclamación del trabajador que sufre daños derivados de la IA y su relación con el régimen general de responsabilidad del empresario.
*Palabras clave:* Inteligencia artificial, seguridad y salud en el trabajo, obligaciones del empresario, obligaciones del proveedor, responsabilidad por daños derivados de IA.

*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:* Artificial Intelligence, employer safety obligations, obligations of provider of AI system, liability rules for Artificial Intelligence.

[*]Este artículo es fruto del Proyecto de investigación "Inteligencia Artificial y Prevención de Riesgos Laborales: retos para la normativa preventiva y en materia de responsabilidad". PID2021-123514NB-I00.

## 1. Seguridad, técnica y riesgo: el modelo regulador en la encrucijada

La necesidad de dar respuesta jurídica a los problemas que plantean los avances científicos constituye un reto fundamental para el Derecho. El legislador ha de afrontar la tarea de traducir en normas jurídicas los estándares de seguridad a los que debe adecuarse el desarrollo de nuevos productos y servicios, teniendo en cuenta los principios éticos que cabe deducir de las normas básicas internacionales y de la Constitución. Este reto de ofrecer un marco jurídico adecuado y de garantizar una cierta seguridad jurídica a los operadores económicos y también a los ciudadanos es una constante que acompaña a las relaciones entre ciencia, técnica y derecho[1].

En el ámbito de la UE este desarrollo de un marco regulador común que garantice la seguridad de los productos resulta, además, imprescindible para asegurar la libre circulación de mercancías. Por ello, desde la UE se ha gestado una prolífica normativa en materia de seguridad del producto, que ha tratado de conjugar dos grandes objetivos: garantizar una cierta seguridad jurídica a los operadores económicos y un elevado nivel de seguridad a los usuarios.

Para conseguir este propósito la UE apostó inicialmente por el conocido como nuevo enfoque en materia de armonización técnica[2], conforme al cual la normativa europea fija los requisitos esenciales de seguridad, que son de cumplimiento obligatorio, y se permite a los organismos de normalización aprobar normas técnicas, que son voluntarias, pero cuyo seguimiento confiere al os productos la presunción de conformidad con los requisitos esenciales contenidos en normas jurídicas obligatorias. Los organismos de normalización son, por tanto, los encargados de la fijación de estándares técnicos a los cuales puede acomodarse el diseño y elaboración de los productos para que se presuman seguros. El sistema se completa con la previsión de un control ex post, que pretende evaluar y acreditar que efectivamente el producto responde a esos estándares y que, en función de la peligrosidad del producto es más o menos exigente. Con este modelo se trata de aproximar las legislaciones de los estados miembros con el fin de garantizar un nivel homogéneo de seguridad de los productos y evitar posibles trabas comerciales a la entrada de productos procedentes de otros países.

Por tanto, en lo referido a la seguridad de los productos la política legislativa de la UE ha pivotado sobre dos principios fundamentales: garantizar, por un lado, un nivel elevado de seguridad sin perjudicar la libre circulación de mercancías; y asegurar, por otro lado, la efectividad de la normativa de seguridad teniendo en cuenta la elevada complejidad técnica de este ámbito normativo. Ciertamente, en el fondo de este modelo regulador se encuentra el reconocimiento de la dificultad de ofrecer una respuesta legal, en forma de norma jurídica obligatoria, que fije los requisitos de seguridad de los productos. Por ello,

[1] ESTEVE PARDO, J.: *Técnica, riesgo y Derecho. Tratamiento del riesgo tecnológico en el Derecho ambiental,* Ariel Derecho, 1999. Del mismo autor: "La regulación de los riesgos: gestionar la incertidumbre", *El Cronista del Estado Social y Democrático de Derecho,* núm. 96-97 (octubre-noviembre), 2021. pp. 32 y ss.

[2] Resolución del Consejo 85/C 136/01, de 7 de mayo de 1985, relativa a una nueva aproximación en materia de armonización y de normalización.

el legislador confía esta labor a los organismos técnicos de normalización, que elaboran las normas técnicas a las que se debe acomodar la fabricación de los productos y presume que los productos que cumplen dichos estándares cumplen también los requisitos esenciales de seguridad, estos sí, previstos en la norma jurídica obligatoria. El sistema se completa con un proceso de auditoría, también confiado a organismos privados de acreditación, que tiene como fin comprobar que efectivamente los productos y equipos se han elaborado conforme a las normas técnicas de seguridad. Este nuevo enfoque en materia de armonización técnica ha inspirado buena parte de la regulación de los equipos y productos de trabajo. Importantes normas como las relativas a máquinas, equipos de protección individual o materiales de construcción, se han basado en este modelo, que trata de garantizar una coordinación adecuada, rápida y eficaz entre normas jurídicas y técnicas. Desde un punto de vista pragmático ha de reconocerse que el modelo ha resultado eficaz; en efecto, a pesar de las suspicacias que generaba al principio, por cuanto partía del propio reconocimiento de la dificultad del ordenamiento jurídico para dar una respuesta eficaz, se ha conseguido dotar a la UE de un marco regulador en materia de seguridad del producto que efectivamente ha garantizado un nivel elevado de seguridad, facilitar la circulación de mercancías y ofrecer seguridad jurídica a los operadores económicos.

Este modelo regulador en materia de seguridad del producto no se mostraba tan eficiente cuando se trataba de garantizar la seguridad de productos, como las sustancias químicas, que podían entrañar riesgos desconocidos. En ocasiones el principal riesgo de algunos desarrollos tecnológicos tiene que ver precisamente con el desconocimiento de su posible potencialidad dañina. Por ello, para este tipo de productos la UE optó por un modelo de regulación que toma en consideración esa incertidumbre respecto a los riesgos que entrañan y que afronta su regulación garantizando el respeto al conocido como principio de precaución. Para ello, el legislador en lugar de confiar en la elaboración de normas técnicas o en la realización de la correspondiente evaluación de riesgos, prefiere someter el producto a una serie de pruebas con el fin de tener información fiable sobre los riesgos que entraña. En este modelo se ha basado, por ejemplo, la regulación en materia de sustancias y productos químicos. El Reglamento conocido como REACH[3], adopta un modelo basado en el riesgo que pueden entrañar determinadas sustancias; en general, se permite la comercialización de sustancias hasta determinada cantidad, para aquellas que se pretenda comercializar en mayores cantidades se requiere su registro y se contempla la necesidad de evaluación por el propio fabricante de determinadas sustancias que se considera que pueden entrañar peligros. Por último, para comercializar determinadas sustancias consideradas peligrosas se requiere la autorización por parte de las autoridades e incluso se permite el establecimiento de restricciones a su comercialización. En términos generales el registro de las sustancias químicas exige recopilar toda la información disponible sobre las características de las mismas y sus posibles riesgos y, si fuera necesario, realizar ensayos para determinar la peligrosidad de la misma y los

[3] El Reglamento (CE) nº 1907/2006, de 18 de diciembre sobre Registro, Evaluación, Autorización y Restricción de sustancias y mezclas químicas.

riesgos que entraña. Constituye, pues, el primer escalón del sistema REACH. A partir de la información recibida se decidirán los procedimientos posteriores de evaluación y la necesidad o no de someter la sustancia a autorización y/o restricciones. El procedimiento instaurado por REACH puede no acabarse con el mero registro de la sustancia, sino que, en ocasiones, al registro le sigue la realización de una evaluación. En este caso es la autoridad competente del Estado en el cual tenga lugar la fabricación o en el que esté establecido el importador la que debe realizar la citada evaluación que puede ser de tres tipos: evaluación de la propuesta de ensayo, evaluación del expediente y evaluación de la sustancia. La evaluación de la sustancia puede dar lugar a que se considere que la sustancia, por la peligrosidad que entraña, debe ser incluida entre las que precisan autorización de la Agencia, o a que se recomiende la restricción de alguno de los usos de la sustancia.

Fuera del ámbito estrictamente referido a la seguridad de los productos, la UE también ha tenido que arbitrar un marco legislativo para prevenir los riesgos en el ámbito laboral, medioambiental o en materia de protección de datos. En estos casos, la UE no ha partido de estándares técnicos más o menos fiables, no se arbitran unas normas concretas cuyo cumplimiento permite presumir la seguridad de un equipo, proceso o actividad, sino que se obliga a la persona que puede generar el riesgo a evaluarlo conforme a unas determinadas directrices y a arbitrar las medidas preventivas oportunas.

Se puede decir, por tanto, que la UE, antes del vertiginoso desarrollo tecnológico que ha conllevado la IA, contaba ya con un cierto bagaje en la regulación de las cuestiones técnicas. Sin embargo, ha de reconocerse que el desarrollo de la inteligencia artificial incorpora elementos nuevos, que trascienden otros avances tecnológicos, por cuanto existe todavía una mayor incertidumbre científica en cuanto a las posibilidades de la inteligencia artificial. Se desconoce su capacidad de aprendizaje, así como cuáles pueden ser las consecuencias de ciertos desarrollos y su impacto a nivel social. Junto a ello, la celeridad con que se suceden los avances en este ámbito dificulta todavía más el desarrollo y aprobación de una respuesta legislativa adecuada, que proporcione un marco jurídico fiable y garantice un nivel de protección elevado.

Estas peculiaridades han llevado a abrir un proceso de reflexión sobre cómo regular este novedoso avance[4]. En el caso europeo este proceso ha tenido como principal objetivo garantizar un marco jurídico que ofreciera cierta seguridad, sin perjudicar la innovación y el necesario desarrollo de esta nueva tecnología con el fin de competir con otros mercados como el chino o estadounidense. Para ello, la UE encargó a un Grupo

[4] Han existido diversas declaraciones sobre los principios éticos más adecuados. Cabe destacar: Principios de Asilomar, disponible en https:// futureoflife.org/open-letter/pause-giant-ai-experiments/, Declaración de Montreal para un desarrollo responsable de la Inteligencia Artificial 2018, , disponible en https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/01/ES-UdeM_Decl-IA-Resp_LA-Declaration_v4.pdf, Declaración sobre Inteligencia artificial robótica y sistemas autónomos, disponible en https://www.bioeticayderecho.ub.edu/archivos/pdf/EGE_inteligencia-artificial.pdf

Vid sobre esta cuestión: LLANO ALONSO, F.H.: "Ética(s) de la inteligencia artificial y derecho consideraciones a propósito de los límites y la contención del desarrollo tecnológico. *Derechos y libertades,* núm. 51, época II, junio 2024, pp. 177 y ss, disponible en https://e-revistas.uc3m.es/index.php/DYL/article/view/8587/6595.

de expertos de alto nivel la elaboración de las Directrices éticas para una IA fiable"[5]. En este documento se apuesta por una IA ética, fiable y robusta y señala que la actuación legislativa debe hacerse depender de las diversas funcionalidades de la IA y su potencial dañoso en función del uso al que se destine. Por otro lado, este documento apunta cuáles serían los métodos para garantizar la fiabilidad de la IA. Entre estos métodos señalaba la necesidad de una evaluación constante y el establecimiento de normas, códigos de conducta, y procesos de normalización y certificación. Combina, por tanto, la evaluación constante de los riesgos con la normalización y certificación. Asimismo, el documento advertía de la necesidad de tener en cuenta que no todas las aplicaciones de la IA entrañan los mismos riesgos.

A partir de estas recomendaciones la Comisión y el Parlamento iniciaron un arduo proceso de diálogo que culminó con la aprobación el 13 de junio por el Parlamento y el Consejo, de la versión final del Reglamento UE 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) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). El texto ha sido finalmente publicado en el DOUE el pasado 12 de julio[6].

## 2. El modelo regulador en materia de IA

El Reglamento, en línea con las propuestas formuladas por el grupo de expertos, parte de un modelo basado en la entidad del riesgo potencial del sistema de IA, atendiendo a sus diversos usos y funcionalidades, pero este modelo se completa con aspectos propios de los otros modelos antes descritos, como el del nuevo enfoque o el de sometimiento a ciertas restricciones o autorizaciones[7].

---

[5] Dirección General de Redes de Comunicación, Contenido y Tecnologías (Comisión Europea): *Directrices éticas para una IA fiable,* 2019, disponible en: https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

[6] Un estudio exhaustivo de esta cuestión en la versión anterior del Reglamento en: GOÑI SEIN, J.L.: "El Reglamento UE de inteligencia artificial y su interrelación con la normativa de seguridad y salud en el trabajo", en AA.VV.: (dir.: EGUSQUIZA BALMASEDA, M.E. y RODRÍGUEZ SANZ DE GALDEANO, B.): *Inteligencia Artificial y Prevención de Riesgos Laborales: Obligaciones y Responsabilidades",* Tirant lo Blanch, Valencia, 2023, pp. 73 y ss. Vid también: RODRÍGUEZ SANZ DE GALDEANO, B.: "Los sistemas de inteligencia artificial en el ámbito laboral y el marco regulador europeo de seguridad del producto" (dir.: EGUSQUIZA BALMASEDA, M.E. y RODRÍGUEZ SANZ DE GALDEANO, B.): *Inteligencia Artificial y Prevención de Riesgos Laborales: Obligaciones y Responsabilidades",* Tirant lo Blanch, Valencia, 2023, pp. 24 y ss, de la misma autora: "Obligaciones del empresario en materia de prevención de riesgos laborales derivadas de la utilización de sistemas de IA", *Revista Galega de Dereito Social,* núm. 18, 2023, pp. 41 y ss.

[7] Señala el profesor MERCADER UGUINA, J.: "El Reglamento de Inteligencia Artificial: frecuentemos el futuro", *Colección Briefs,* 20/3/2024, AEDTSS, disponible en: https://www.aedtss.com/el-reglamento-de-inteligencia-artificial-frecuentemos-el-futuro/, que el nuevo Reglamento se basa en el triángulo de oro:

El objetivo reconocido es crear un marco jurídico que garantice un nivel de seguridad elevado, que sea coherente con el resto de normativa, que dé seguridad a los operadores jurídicos, sin perjudicar el desarrollo de la IA y el posicionamiento de Europa en la carrera internacional para desarrollar y explotar la IA. En coherencia con ello, el Reglamento parte de un modelo de gestión del riesgo, que modula las exigencias de seguridad en función de la potencialidad dañosa del sistema de IA. Para ello, la IA lleva una clasificación de los sistemas de IA atendiendo a sus funcionalidades y al ámbito o sector en el que se desarrollan. De acuerdo con ello, existirían una serie de sistemas de IA prohibidos, por entrañar un riesgo inaceptable, en un segundo nivel estaría la IA de alto riesgo, y en los dos últimos niveles se encontrarían los sistemas que deben observar determinadas obligaciones de transparencia debido a que interactúan con personas y los sistemas que no entrañan riesgo o simplemente un riesgo mínimo[8]. Junto a ello, la última versión del Reglamento ha incorporado prescripciones específicas para los modelos de IA de propósito general, entre los que se incluirían sistemas que han tenido un desarrollo fulminante en los últimos años, como Chat Gpt.

A partir de aquí, el grueso del Reglamento se centra en regular los requisitos de seguridad que deben reunir los sistemas de IA de alto riesgo. En esta categoría se incluyen dos tipos de sistemas de IA, por un lado, aquellos sistemas que se utilizan en determinados ámbitos importantes y cuyo funcionamiento puede afectar a la salud, seguridad o derechos fundamentales. En lo que aquí interesa, entre estos sistemas se encontrarían, de acuerdo con lo dispuesto en el apartado 4 del Anexo III, los destinados a utilizarse en el empleo, gestión de trabajadores y acceso al autoempleo, en concreto los sistemas destinados a la contratación o selección de personas o los utilizados para tomar decisiones que afecten a las condiciones de las relaciones de índole laboral[9].

Por otro lado, se incluyen entre los sistemas de IA de alto riesgo aquellos que están destinados a incorporarse a productos o ser un componente de seguridad de un producto que ya está regulado por la normativa de armonización técnica de la UE; como, por ejemplo, los sistemas de IA destinados a integrarse en una máquina o ser un componente de seguridad de dicha máquina.

Para estos sistemas de IA de alto riesgo el Reglamento contempla la necesidad de cumplir unos determinados requisitos de seguridad, previstos en la sección 2 del capítu-

---

aproximación desde el riesgo, garantías y responsabilidades. ESPÍN ALBA, I.: "Sesgos discriminatorios en la toma automatizada de decisiones en la contratación y protección de datos" en AA.VV.: *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* Aranzadi, Pamplona, 2024, pp. 127 y ss, advierte el paralelismo de la regulación en materia de IA con la relativa a protección de datos, sobre todo, en lo que se refiere a la exigencia de transparencia.

[8] ARTIÑANO MARRA, P. y SÁNCHEZ ORO, J.: "Responsabilidad por el funcionamiento de sistemas de inteligencia artificial: los desafíos de la "documentación técnica del Reglamento de Inteligencia Artificial", *Derecho Digital e Innovación,* núm. 20, 2024. FERNÁNDEZ HERNÁNDEZ, C.: "El Reglamento de Inteligencia Artificial. Un nuevo marco regulatorio para una tecnología en continua evolución", *Derecho Digital e Innovación,* núm. 19, 2024.

[9] LLORENS ESPADA, J.: "La inteligencia artificial para la mejora de la seguridad y salud laboral y su encaje en el marco regulatorio europeo", *Trabajo y Derecho,* núm. 19, 2024.

lo III. Así, el art. 8 comienza señalando que los sistemas de alto riesgo habrán de cumplir los requisitos recogidos en la sección teniendo en cuenta la técnica y su funcionalidad. Para garantizar el cumplimiento de estos requisitos se tendrá en cuenta el sistema de gestión de riesgos contemplado en el art. 9. Este artículo, que bebe de los modelos normativos basados en la gestión del riesgo, señala que se deberá implantar un sistema de gestión de riesgos que constará de las siguientes etapas: determinación de riesgos y análisis, estimación y evaluación de riesgos tanto derivados de un uso previsto como de un uso indebido razonablemente previsible, evaluación de otros riesgos que puedan surgir a partir del análisis de datos, adopción de medidas adecuadas para hacer frente a los riesgos detectados. Se señala, por último, que los sistemas de IA de alto riesgo serán sometidos a pruebas para determinar cuáles son las medidas de gestión de riesgo más adecuadas, entre estos procedimientos de prueba cabe la realización de pruebas en condiciones reales. Además, la sección 2 contempla también una serie de obligaciones para aquellos sistemas de IA que utilicen entrenamiento con datos.

Esta obligación de establecer un sistema de gestión de riesgos, se completa con una serie de obligaciones relativas al aseguramiento de la calidad de los datos, conservación de registros, garantizar la transparencia y la correcta información a los responsables del despliegue, establecer medidas de supervisión humana y garantizar la precisión, solidez y ciberseguridad (artículos 10 a 15).

Se observa, por tanto, cómo esta sección 2 recoge lo que en la normativa basada en el nuevo enfoque serían los requisitos generales de seguridad; de entre estas obligaciones la principal para el proveedor es el establecimiento de un sistema de gestión de riesgos, que garantice que el sistema no conlleva riesgos inaceptables y que contemple las medidas preventivas adecuadas. En este punto, la normativa de IA se separa de la normativa tradicional de nuevo enfoque puesto que a diferencia de esta normativa no establece las orientaciones genéricas de esas medidas preventivas. Por ejemplo, el reglamento máquinas, contiene en su larguísimo Anexo III, toda una serie de prescripciones que fijan los requisitos de seguridad de aspectos referidos a resguardos, ruido, riesgo de incendio, dispositivos de parada, etc.

Una vez cumplidos estos requisitos de seguridad previstos en la sección 2, el proveedor de un sistema de IA de alto riesgo debe evaluar la conformidad del sistema con tales requisitos. Nuevamente el RIA entronca con el espíritu propio de la normativa de nuevo enfoque que, según se ha comentado, fija requisitos de seguridad genéricos y confia en agentes privados y en el propio fabricante la evaluación de la conformidad a dichos requisitos. En el caso del RIA, se ha previsto que cada Estado designe una autoridad notificante que será la responsable de autorizar a los organismos que se encargarán de evaluar la conformidad de los equipos.

En cuanto a los procedimientos de evaluación se recogen en el art. 43 del RIA y varían en cuanto a su exigencia dependiendo del tipo de sistema de alto riesgo y de la observación de normas armonizadas y especificaciones comunes. Aunque resulte un tanto complejo vale la pena ordenar estos sistemas de evaluación de conformidad, aunque sea muy someramente.

a) Sistemas de IA de alto riesgo del Anexo III.1 utilizados en biometría.
b) Sistemas de IA de alto riesgo del Anexo III apartados 2 a 8. Entre los que se incluye los destinados al empleo, gestión de los trabajadores y autoempleo.
c) Sistemas de IA que forman parte de productos ya sometidos actos legislativos de armonización de la UE recogidos en el Anexo I del RIA (entre ellos, directiva máquinas, equipos a presión, equipos de protección individual…).

Con el fin de evitar duplicidades, para aquellos sistemas de IA destinados a formar parte de equipos o productos que ya disponen de su propia normativa técnica, el RIA ha previsto que los proveedores puedan integrar los procesos de prueba y presentación de información de acuerdo con los procedimientos ya previstos en la legislación de armonización sectorial y específica.

En este punto, entra en juego un pilar esencial de la normativa de nuevo enfoque que es el referido a las normas armonizadas. Tal y como se ha señalado, en el esquema general del nuevo enfoque en materia de armonización técnica las normas técnicas elaboradas por el organismo de normalización adquieren una importancia trascendental, por cuanto su cumplimiento por el fabricante supone que se presume automáticamente la conformidad del producto con los requisitos obligatorios. En el caso del RIA el art. 40 recoge expresamente esta presunción. Ahora bien, en el caso de la IA el problema fundamental es la falta de normas técnicas. A diferencia de lo que ocurre en otros ámbitos como el de máquinas, en el que existen ya normas técnicas plenamente consensuadas por los organismos de normalización y que garantizan un nivel elevado de seguridad, en el caso de la IA no existe un consenso claro en cuanto a estas soluciones técnicas. Así se refleja en el propio RIA, cuando en su art. 40.2 señala que la Comisión formulará sin demora peticiones de normalización que contemplen los requisitos de seguridad exigidos por el Reglamento. Además, el art. 41 contempla la posibilidad de que la Comisión pueda directamente adoptar especificaciones comunes para los requisitos de seguridad de la sección 2; esta posibilidad de que sea la Comisión la que adopte las especificaciones comunes se condiciona a la falta de existencia de normas armonizadas.

Esta necesidad de que se desarrollen guías y códigos de buenas prácticas ha conducido a la aprobación del RD 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial. Este RD pretende crear lo que se conoce como un *sandbox* en el que puedan participar las empresas que lo soliciten con el fin de efectuar pruebas para examinar la forma de poner en práctica los requisitos de seguridad aplicables a los sistemas de IA, el objetivo es generar guías basadas en la evidencia, que ayuden a las empresas al cumplimiento de los requisitos exigidos por la nueva normativa de IA.

Todas las obligaciones descritas hasta el momento afectan únicamente a los sistemas calificados como de alto riesgo. Los proveedores del resto de sistemas de IA simplemente han de observar obligaciones de transparencia o información, dirigidas a informar a los usuarios de que están interactuando con un sistema de IA.

## 3. Obligaciones en materia de prevención de riesgos laborales

### *3.1. Obligaciones del fabricante de equipos de trabajo*

En lo que se refiere a las relaciones entre el bloque normativo relativo a la seguridad del producto y el relativo a la seguridad y salud en el trabajo, ha de tenerse en cuenta que la obligación del fabricante de sistemas de IA no desplaza las obligaciones específicas en materia de seguridad y salud en el trabajo[10]. El art. 2.11 RIA señala que el Reglamento no impedirá que la Unión o los Estados mantengan o introduzcan disposiciones más favorables a los trabajadores o fomenten la aplicación de convenios colectivos más favorables. De acuerdo con ello, el RIA no desplaza la íntegra aplicación de la LPRL y, en particular, lo dispuesto en el art. 41 LPRL, que recoge el régimen general de obligaciones del empresario y fabricante respecto de los equipos de trabajo.

El artículo se divide en dos apartados, el primero recoge el elenco de obligaciones de los fabricantes, importadores y suministradores de maquinaria, equipos y productos de trabajo. Con carácter general señala que están obligados asegurar que no constituyen una fuente de peligro, siempre que sean instalados en la forma recomendada. Asimismo, obliga a los fabricantes de sustancias químicas a envasarlas y etiquetarlas de forma adecuada. Obliga también a suministrar información sobre la forma correcta de uso, medidas preventivas adicionales y riesgos de un uso normal o empleo inadecuado. Por último, los fabricantes de EPIs están obligados a garantizar su efectividad[11].

El artículo se cierra con un apartado 2 que obliga al empresario a garantizar que la información sobre la utilización, manipulación y riesgos que entrañan la maquinaria equipos y productos se faciliten a los trabajadores en términos comprensibles.

Se trata, como puede observarse, de una obligación ciertamente genérica y exigente; los fabricantes deben garantizar que los equipos no constituyan una fuente de peligro e informar sobre sus riesgos residuales y medidas preventivas recomendadas. Parece, por tanto, que la LPRL maneja un concepto de riesgo diferente al de la normativa de seguridad del producto y al de la normativa específica sobre riesgos de IA; puesto que el objetivo de la LPRL es asegurar la ausencia de peligro en los productos[12]. En la interpretación de la LPRL y en la concreción de estos conceptos, ciertamente indeterminados, en torno a la ausencia de riesgo y medidas preventivas han de tenerse en cuenta los propios principios preventivos recogidos en el art. 15 LPRL que obligan a evitar los riesgos en el origen, evaluar los que no se puedan evitar, sustituir lo peligroso por lo que entrañe poco o ningún peligro, a tener en cuenta la evolución de la técnica y a planificar las medidas

[10] RODRÍGUEZ SANZ DE GALDEANO, B.: "La responsabilidad empresarial por accidentes vinculados a la inteligencia artificial", *Trabajo y Derecho: nueva revista de actualidad y relaciones laborales,* núm. extra 19, 2024.

[11] RODRÍGUEZ SANZ DE GALDEANO, B.: *Las responsabilidades de los fabricantes en materia de prevención de riesgos laborales,* Lex Nova, Valladolid, 2005.

[12] GOÑI SEIN, J.L.: "Sistemas de inteligencia artificial y prevención de los riesgos laborales: obligaciones del proveedor y del empresario", *Labos*, (en este nº monográfico).

preventivas adaptando el trabajo a la persona y adoptando medias de protección colectiva e individual.

Por lo tanto, la LPRL asume un cierto nivel de riesgo en los productos, así lo pone de manifiesto el que se refiera a la adopción de medidas preventivas y de protección frente a riesgos; ahora bien, el objetivo final de la LPRL es el de ofrecer un producto seguro desde el diseño y que, en caso de que subsistan riesgos, sean controlados. Parece, por tanto, que el modelo de seguridad buscado es un modelo que combina la seguridad en el diseño del producto con la necesaria adopción de medidas complementarias de carácter preventivo, con el fin de minimizar los riesgos y proteger a los trabajadores frente a los riesgos residuales.

El precipitado de todo lo anterior es que el cumplimiento de la normativa de seguridad del producto y de la normativa específica en materia de IA no garantiza que el fabricante esté cumpliendo con su obligación de asegurar equipos y productos de trabajo que no constituyan una fuente de peligro. La razón principal de ello se encuentra, como se ha viene insistiendo, en el diferente modelo de riesgo y seguridad.

Este diferente rasero en la valoración de la seguridad, se ha visto acrecentado tras la aprobación del RIA. Tal y como se ha visto, el RIA pretende solo establecer requisitos de seguridad para algunos sistemas de IA considerados como de alto riesgo. Para el resto de sistemas no se prevén obligaciones de seguridad o son mucho más livianas.

Junto a ello ha de tenerse en cuenta que el propio modelo de seguridad del producto puede entrañar fugas que determinen que un producto, incluso contando con el marcado CE y la correspondiente declaración CE, no sea conforme con los requisitos esenciales de seguridad. Así puede suceder cuando no se realiza correctamente la evaluación de conformidad y se consideran conformes productos que no lo son. Así puede también acontecer cuando, aunque se hayan seguido las normas técnicas armonizadas, estas no respondan a los estándares de seguridad exigidos obligatoriamente.

En el caso de los sistemas de IA, estas posibles fugas pueden aumentar debido a la incertidumbre en cuanto a cuáles deban ser los estándares técnicos que permitan presumir la conformidad con las obligaciones generales. Todavía no se disponen de estas normas técnicas y se trata de un ámbito en el que los organismos de normalización tradicionales no cuentan todavía con especialización suficiente.

### *3.2. Obligaciones del empresario que incorpora sistemas de IA*

En cuanto a las obligaciones del empresario, el RIA no desplaza la aplicación de la LPRL; y a sus obligaciones como responsable del despliegue se suman sus obligaciones específicas en materia preventiva. Al respecto, el RIA señala expresamente que las obligaciones del responsable del despliegue no afectan a otras obligaciones que el Derecho nacional imponga (art. 26). En consecuencia, el empresario es el máximo garante de la seguridad y salud de los trabajadores y a sus deberes preventivos se suman las obligaciones derivadas del nuevo RIA, en cuanto se convierte en responsable del despliegue del sistema de IA y también, si realiza modificaciones, puede llegar a tener las obligaciones propias del proveedor.)

El art. 3.4 se considera responsable del despliegue a la "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".

Según el art. 26, el empresario, en cuanto responsable del despliegue, debe:

a) Adoptar medidas técnicas y organizativas para garantizar que los sistemas de IA de alto riesgo se usan con arreglo a las instrucciones.
b) Encomendar la supervisión humana a personas físicas con competencia, formación y autoridad necesarias. De acuerdo con el art. 14 es necesaria.
c) Vigilar el funcionamiento del sistema conforme a las instrucciones de uso e informar cuando detecten que pueden entrañar un riesgo o existe un incidente.
d) Conservar los archivos de registro generados que estén bajo su control.
e) Informar a los representantes de los trabajadores y trabajadores afectados.
f) Utilizar la información facilitada para realizar la evaluación de impacto.
g) Informar a las personas físicas de que están expuestas a sistemas de alto riesgo que pueden tomar decisiones respecto de ellas.

A estas obligaciones específicas que les impone el RIA, se suman como se ha mencionado, las derivadas de la LPRL. De acuerdo con ello, el empresario sigue siendo el principal obligado en materia de seguridad y salud en el trabajo y debe adquirir productos seguros, realizar la evaluación de riesgos de todos los sistemas de IA, no solo los de alto riesgo, y adoptar las medidas preventivas necesarias. Además, de acuerdo con el art. 41 LPRL, se refuerzan las obligaciones de uso conforme de los sistemas, sean o no de alto riesgo, y el deber de instalarlos y mantenerlos en las condiciones descritas por el fabricante, así como de informar sobre los riesgos que entrañan.

En lo que se refiere a la obligación del empresario de adquirir productos seguros ha de advertirse ha de advertirse nuevamente que la adquisición de un producto conforme con la normativa de seguridad, no significa que en materia preventiva tanto el fabricante como el empresario se encuentren libres de responsabilidad. Puede ocurrir que las obligaciones de seguridad del producto no se correspondan con las más específicas referidas a la seguridad en el trabajo y prevención de riesgos de los usuarios profesionales.

## 4. Marco normativo en materia de responsabilidad por daños derivados de la IA: complementariedad de los sistemas normativos en materia de responsabilidad

### *4.1. Las iniciativas legislativas en el ámbito de la UE para la armonización del régimen de responsabilidad por daños derivados de la IA*

Junto con la regulación de los requisitos de seguridad de los sistemas de IA, el otro gran frente regulador para la UE ha sido el relativo a la responsabilidad. Se veía que era necesario acomodar la normativa existente a los nuevos desafíos que plantea la IA.

En este ámbito, el legislador comunitario ha optado por establecer un régimen unificado para los daños causados por productos defectuosos y adaptar la regulación existente a los nuevos desafíos derivados de la IA[13]. Para ello, ha aprobado una nueva Directiva que deroga a la anterior en materia de responsabilidad por productos defectuosos[14].

Paralelamente, el legislador europeo ha trabajado en una propuesta de Directiva específica para los daños derivados de IA[15]. La propuesta no trata de unificar los regímenes de responsabilidad vigentes, sino simplemente de establecer una serie de facilidades para la prueba[16]. Con este objetivo, la propuesta de Directiva se basa en el incumplimiento del deber de diligencia, pero establece, por un lado, mecanismos para facilitar el acceso del perjudicado a las pruebas y, por otro lado, fija una serie de presunciones refutables en cuanto a la observación del deber de diligencia[17]. En este punto la propuesta presenta una novedad importante y es que trata de dibujar un sistema coherente con el previsto para abordar la seguridad de los productos. Por ello, un elemento central para probar la diligencia es la observación de los requisitos de seguridad, de tal manera que los proveedores que acrediten que han cumplido las prescripciones del RIA podrán liberarse de responsabilidad.

No obstante, por el momento, esta propuesta de Directiva no ha visto la luz y parece que atendiendo al principio de subsidiariedad y una vez adoptada la reforma de la Directiva de responsabilidad por productos defectuosos, su aprobación no constituye una prioridad para la UE.

En todo caso, en su redacción actual, la propuesta de Directiva sobre responsabilidad por daños derivados de la IA se proyecta como una normativa complementaria,

---

[13] Vid ampliamente JORQUI AZOFRA, M.: *Responsabilidad por daños causados por productos y sistemas de inteligencia artificial,* Dykinson, Madrid, 2023. Vid también al respecto: NAVAS NAVARRO, S.: "Régimen europeo en ciernes en materia de responsabilidad deriva de los sistemas de Inteligencia artificial", *Revista CESCO de Derecho del Consumo,* núm. 44, 2022, pp. 43 y ss. EBERS, M.: "La utilización de los agentes electrónicos inteligentes en el tráfico jurídico. ¿Necesitamos reglas especiales en el Derecho de la responsabilidad civil?", *Indret,* núm. 3, 2016; ATIENZA NAVARRO, M.L.: *Daños causados por inteligencia artificial y responsabilidad civil,* Atelier, Barcelona, 2022; del mismo autor: "La responsabilidad civil por daños causados por inteligencia artificial. Estado de la cuestión" en AA.VV.: *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* Aranzadi, Pamplona, 2024, pp. 341 y ss. ATAZ LÓPEZ, J.: "Daños causados por las cosas una nueva visión a raíz de la robótica y de la inteligencia artificial", *Working Paper*, 4/2020, UB, disponible en: http://hdl.handle.net/2445/169850; NAVAS NAVARRO, S.: *Daños ocasionados por sistemas de inteligencia artificial. Especial atención a su futura regulación,* Comares, Granada, 2022; DE SILVA LÓPEZ DE LETONA, J.: "Responsabilidad por daños causados por sistemas de inteligencia artificial", *Derecho Digital e Innovación,* núm. 11, 2022.

[14] Aprobada por el Parlamento Europeo y el Consejo el 12 de marzo de 2024.

[15] 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, de 28 de septiembre de 2022.

[16] PEÑA LÓPEZ, F.: "Responsabilidad objetiva y subjetiva en las propuestas legislativas europeas sobre responsabilidad civil aplicables a la inteligencia artificial" en AA.VV.: *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* Aranzadi, Pamplona, pp. 411 y ss.

[17] EVANGELIO LLORCA, R.: "Causalidad y responsabilidad civil por daños ocasionados por sistemas de inteligencia artificial: las presunciones de causalidad en las propuestas normativas de la UE", en AA.VV.: *Derecho de contratos…,* op. cit., pp. 596 y ss.

que no desplaza a la Directiva de productos defectuosos. La diferencia fundamental entre ambos regímenes de responsabilidad, el de la Directiva de productos defectuoso y el de la propuesta de responsabilidad civil extracontractual, se encuentra en que la normativa de productos defectuosos parte de un sistema de responsabilidad objetivo, aunque contempla una serie de causas de exoneración, y limita los daños susceptibles de ser reclamados conforme a la misma. Por su parte, la normativa de responsabilidad civil por daños derivados de la IA, no pretende armonizar los sistemas de responsabilidad civil de los Estados, que pueden mantener su régimen subjetivo u objetivo, sino simplemente facilitar al perjudicado la reclamación por los daños derivados de estos sistemas, teniendo en cuenta las particularidades de su funcionamiento (opacidad, aprendizaje, etc)[18].

### *4.2. Ámbito de aplicación de la Directiva de productos defectuosos y relación con otros regímenes de responsabilidad*

En lo que se refiere a su ámbito de aplicación, la nueva Directiva sobre responsabilidad por productos defectuosos señala en el art. 3 que los Estados miembros no introducirán disposiciones más o menos estrictas que lleven al establecimiento de un diferente nivel de protección de los consumidores y personas físicas. Por lo tanto, en materia de responsabilidad por daños causados por productos defectuosos, el legislador de la Unión ha optado por fijar un régimen uniforme, sin margen para que los Estados establezcan disposiciones más rigurosas. Ahora bien, el art. 2.4 apartado b), relativo al ámbito de aplicación, aclara que la Directiva no afecta a los derechos que la persona perjudicada tenga en virtud de las normas nacionales en materia de responsabilidad contractual o extracontractual por motivos distintos del carácter defectuoso de un producto. En relación con ello el considerando 9 aclara que: "una persona perjudicada podría reclamar una indemnización por daños y perjuicios sobre la base de la responsabilidad contractual o por motivos de responsabilidad extracontractual que no se refieran a la responsabilidad del fabricante por el carácter defectuoso de un producto" y añade: "Esto afecta, por ejemplo, a la responsabilidad basada en una garantía o en la culpa…".

Por su parte, el art. 6 recoge los daños indemnizables en virtud del régimen de responsabilidad por productos defectuosos[19]. Ha de tenerse en cuenta que esta limitación en cuanto a los daños indemnizables conforme al régimen específico de la Directiva no impide que se reclamen los daños no incluidos en virtud de otros regímenes de responsabilidad, así lo aclara el apartado 3[20].

[18] Vid. al respecto: EGUSQUIZA BALMASEDA, M.E.: "Marco normativo general y propuestas de regulación en la responsabilidad civil" en AA.VV. (dir.: RODRÍGUEZ SANZ DE GALDEANO, B. y EGUSQUIZA BALMASEDA, M.E.): *Inteligencia Artificial…,* op. cit., pp. 24 y ss

[19] RUBÍ PUIG, A.: "Inteligencia artificial y daños indemnizables", en AA.VV.: *Derecho de contratos, responsabilidad…,* op. cit., pp. 622 y ss.

[20] Sobre la necesidad de prever reglas particulares cuando se trate de uso profesional vid: VAZQUEZ DE CASTRO, E.: "Aproximación a la responsabilidad derivada de los riesgos de la inteligencia artificial en Eu-

Entre estos daños susceptibles de ser indemnizados se encuentran la muerte o lesiones, incluidos los daños a la salud psicológica, se incluyen también los daños morales siempre que puedan ser indemnizados de acuerdo con el derecho nacional. También se incluyen los daños derivados de la destrucción de cualquier propiedad, salvo los causados en el propio producto defectuoso o en las propiedades utilizadas para fines profesionales, los datos utilizados con fines profesionales. Ha de advertirse que esta exclusión afecta a los daños y no impide, por tanto, que el trabajador, a pesar de ser un usuario profesional y no un consumidor en sentido estricto, pueda reclamar conforme al régimen previsto en la nueva Directiva[21].

En suma, el trabajador podría reclamar los posibles daños del proveedor si media negligencia, conforme a lo previsto en el régimen de responsabilidad extracontractual del art. 1902 Cc.; asimismo, el trabajador podría reclamar del empresario los posibles daños por el incumplimiento de sus obligaciones preventivas o de sus obligaciones específicas en cuanto responsable del despliegue de un sistema de IA.

Por su parte, el empresario podría también dirigirse frente al fabricante o proveedor y exigirles responsabilidad contractual, si ha mediado un contrato entre ambos o responsabilidad extracontractual si ha existido negligencia y consigue acreditarla.

### *4.3. La responsabilidad por daños causados por sistemas de IA de acuerdo con la nueva Directiva de productos defectuosos*

#### 4.3.1. El concepto de producto

Una de las principales novedades de la Directiva es la adaptación de la definición de producto para incorporar los nuevos desarrollos derivados de la IA[22]. Así, el art. 4.1 define producto como cualquier bien mueble, aun cuando esté incorporado a otro bien mueble o a un bien inmueble o interconectado con estos; también la electricidad, los archivos de fabricación digital, las materias primas y los programas informáticos. A continuación, define archivo de fabricación digital como una versión digital o plantilla digital que contiene información.

De acuerdo con lo anterior estarían incluidos en el ámbito de aplicación de la Directiva los programas informáticos y los archivos de fabricación digital que puedan

ropea", AA.VV. (SOLAR CAYÓN, J.I.): *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho,* Editorial Universidad de Alcalá, 2020, pp. 239 y ss

[21] Vid al respecto: JORQUI AZOFRA, M.: "Responsabilidad por los daños…", op. cit., pp. 280; también lo ponía de manifiesto conforme al régimen de la anterior Directiva, PARRA LUCÁN, M.A.: "Responsabilidad civil por productos defectuosos" en REGLERO CAMPOS, L.F. (COORD.): *Tratado de Responsabilidad civil,* Aranzadi, Pamplona, 2014.

[22] JORQUI AZOFRA, M.: "El concepto legal de producto a la luz de la nueva propuesta de directiva sobre responsabilidad por los daños causados pro productos defectuosos", en AA.VV. (dir.: RODRÍGUEZ SANZ DE GALDEANO, B. y EGUSQUIZA BALMASEDA, M.E.): *Inteligencia Artificial y prevención de riesgos laborales: obligaciones y responsabilidades,* Tirant lo Blanch, Valencia, 2023, pp.331 y ss.

funcionar de forma independiente. También, cuando estos archivos se incorporan a un producto, sería, tal y como señala el considerando 16, el caso de aquellos archivos que permiten el control automatizado de máquinas herramientas.

Igualmente, la Directiva se ha preocupado por incluir los servicios conexos entendidos como servicios integrados en un producto o interconectados con él. Estos servicios se consideran componentes del producto siempre que estén bajo el control del fabricante. Sería el supuesto, tal y como aclara el considerando 17, de servicios digitales integrados o conectados con el producto para que pueda funcionar.

La cuestión fundamental en estos casos en los que el archivo de fabricación digital o el servicio conexo se integran en un producto, será determinar quién es el sujeto responsable por un posible defecto. Puede suceder que los daños se deriven de un mal funcionamiento del archivo o servicio, o de una falta de actualización, o de la realización de mejoras etc. La respuesta dependerá, tal y como se verá posteriormente, de quién tenga el control sobre el producto y de si eran exigibles determinadas actualizaciones o mejoras al fabricante del componente (archivo digital o servicio conexo).

### 4.3.2. El concepto de producto defectuoso

La prueba del defecto del producto se erigía en la anterior Directiva en el elemento determinante para la imputación de responsabilidad[23]. La nueva Directiva ha mantenido también la prueba del defecto del producto en la base del sistema de responsabilidad, sin embargo, ha introducido nuevos elementos para caracterizar un producto como defectuoso y ha previsto una serie de facilidades en materia probatoria acordes con la complejidad técnica de algunos productos y, muy en particular, de los nuevos sistemas de IA. Partiendo de este esquema, el defecto del producto y su prueba se convierten en los dos elementos centrales para la imputación de responsabilidad.

El art. 7 contiene la definición de producto defectuoso y señala que se considerará defectuoso un producto cuando no ofrezca la seguridad que una persona tiene derecho a esperar y que se exige asimismo en virtud del Derecho de la Unión o nacional. En el apartado 2 se incluyen una serie de circunstancias que habrán de ser valoradas. Este listado de circunstancias se ha actualizado con el fin de incorporar algunas particularidades propias del funcionamiento de los sistemas de IA. Así, se señala que habrá de tomarse en consideración la posibilidad de que el producto siga aprendiendo o adquiera nuevas características, el efecto previsible de otros productos que se utilicen a la vez o estén interconectados y también apunta que un producto no se considerará defectuoso por la única razón de que se introduzcan en el mercado productos mejores, incluidas la as actualizaciones o mejoras.

En lo que aquí interesa se incluye también la necesidad de tener en cuenta el uso indebido pero razonable; al respecto, el considerando 31 de la Directiva, a modo de

[23] EVANGELIO LLORCA, R.: "Causalidad y responsabilidad civil por daños ocasionados por sistemas de inteligencia artificial: las presunciones de causalidad en las propuestas normativas de la UE", en AA.VV.: *Derecho de contratos…,* op. cit., pp. 577 y ss.

ejemplo, alude expresamente al comportamiento previsible de un usuario de maquinaria derivada de una falta de concentración. Por lo tanto, habrá de considerarse las condiciones propias de un uso profesional (tareas repetitivas, presencia de otros usuarios, sometimiento a exigencias de tiempo, etc.)

En la valoración de la seguridad del producto se tiene en consideración también los requisitos de seguridad pertinentes incluidos los requisitos de ciberseguridad. En este punto la Directiva conecta con la regulación sobre seguridad del producto, incluyendo como un elemento más a la hora de valorar la seguridad del producto los requisitos de seguridad definidos en dicho bloque normativo. Igualmente, contempla que habrán de tomarse en consideración las intervenciones de las autoridades componentes entre las que incluye la retirada de productos, no obstante, también señala que dichas intervenciones por sí solas no suponen la defectuosidad del producto (art. 7.2, g y considerando 34). Sin embargo, más allá de esta disposición, ha de tenerse en cuenta que la Directiva de productos defectuosos no traslada el sistema de pirámide de riesgos del Reglamento que, en cambio, sí que ha sido tenido en cuenta por la propuesta de Directiva de Responsabilidad en materia de IA. En el caso de la Directiva de productos defectuosos, lo esencial es determinar el defecto del producto atendiendo a los factores apuntados, con independencia de que el sistema sea calificado como de alto riesgo o no a efectos del Reglamento de IA.

A la hora de valorar la defectuosidad del producto deviene esencial el momento de la comercialización o puesta en servicio y el control por parte del fabricante. En este aspecto la Directiva ha debido adaptarse a las características propias de esta nueva tecnología, capaz de seguir evolucionando. En el art. 11 se contempla como causa de exención de responsabilidad que sea probable que el defecto no existía en el momento en que el producto se introdujo en el mercado o fue puesto en servicio; ahora bien, esta causa de exención tiene una excepción referida precisamente a la existencia de mejoras o actualizaciones necesarias para el funcionamiento del producto siempre que estén bajo el control del fabricante. De acuerdo con las definiciones de la Directiva se entiende que el fabricante tiene el control cuando autoriza la integración, interconexión suministro de componentes, que incluyen actualizaciones o mejoras de los programas, cuando tiene capacidad de suministrar actualizaciones o mejoras de programas informáticos o cuando modifica el producto. Al respecto, el considerando 19 señala que*:* "una vez que un producto se haya introducido en el mercado debe considerarse que permanece bajo el control del fabricante siempre que este conserve la capacidad de suministrar actualizaciones o mejoras de los programas informáticos". La valoración de las consecuencias que la pérdida de control puede tener en la consideración del producto como defectuoso ha de ponerse en relación con lo dispuesto en el apartado del art. 7, que señala que ha de tenerse en cuenta la posibilidad de que el producto siga aprendiendo. Parece, por tanto, que el fabricante no puede evitar la necesidad de advertir cuáles son las necesidades de seguimiento del producto y, en su caso, realizar las mejoras y actualizaciones que resulten oportunas.

Con todo, el carácter defectuoso del producto no implica automáticamente responsabilidad del proveedor ya que el art. 11 contempla una serie de causas de exención de responsabilidad.

Por un lado, se encontrarían una serie de causas que tienen que ver con la falta de comercialización o puesta en servicio del producto o con la falta de existencia del defecto en el momento de la comercialización (apartados a, b y c).

Por otro lado, se encontraría aquellos casos en que el defecto se debe al cumplimiento de normas obligatorias. Se trata de un supuesto poco frecuente, que requiere que el defecto venga motivado precisamente por la observancia de normas que han de ser obligatorias; no parece que quepa entender como normas obligatorias las normas técnicas o códigos de conducta.

La Directiva ha mantenido como causa de exoneración los denominados riesgos del desarrollo, esto es, los casos en los que el estado de los conocimientos científicos no permitía descubrir el carácter defectuoso del producto en el momento de su introducción en el mercado o puesta en servicio, o mientras estuvo bajo el control del fabricante.

Por último, se contempla una batería de causas de exoneración referidas a que el defecto no es imputable al fabricante y a las posibles consecuencias de la actuación de otras personas en la consideración del producto como defectuoso. Así, el art. 11 f) señala que el fabricante de un componente defectuoso se puede exonerar siempre que pruebe que el defecto del producto se debe al diseño o a las instrucciones dadas por el fabricante del producto. El apartado g) señala que, cuando una persona modifique de forma sustancial el producto, se exonerará de responsabilidad si el carácter defectuoso no se debe a la modificación. Esta exención ha de ser leída en coherencia con el art. 8.2 que señala que cualquier persona que modifique un producto se convierte en fabricante, siempre que la modificación sea sustancial y, en consecuencia, asumirá la responsabilidad por los daños que cause, salvo que entre en juego la exoneración del apartado g), que se acaba de comentar. Ahora bien, el apartado 2 del art. 11 precisa que no cabrá exoneración de responsabilidad cuando el producto esté bajo el control del fabricante o de un operador económico y el defecto se deba a un servicio conexo, a programas informáticos, actualizaciones o mejoras, a la falta de actualizaciones o mejoras de los programas informáticos o a una modificación sustancial. Dicho de otro modo, cuando el fabricante tenga el control del producto y realice mejoras será responsable. Asimismo, serán equiparados al proveedor y considerados como responsables de los posibles daños quienes realicen modificaciones sustanciales, siempre que el defecto tenga su origen en tales modificaciones.

### 4.3.3. La carga de la prueba

Una de las principales novedades de la Directiva es la introducción de una serie de presunciones con el fin de facilitar la prueba a la persona perjudicada. La Directiva ha sido consciente de que la complejidad técnica de algunos productos puede dificultar la prueba del defecto. Además, en la fijación del elenco de presunciones dirigido a facilitar la carga del a prueba la Directiva ha tratado de mantener un régimen coherente con el RIA.

Para ello, en primer lugar, trata de ayudar al perjudicado a acceder a las pruebas que le permitan demostrar el carácter defectuoso del producto. Así el art. 9 señala que los

estados han de garantizar a los órganos jurisdiccionales nacionales las facultades necesarias para que, a petición de una persona perjudicada, puedan requerir los hechos y pruebas que ayuden a respaldar la reclamación de daños, siempre que resulte proporcional.

En segundo lugar, recoge en su art. 10 una serie de presunciones, que se explicarán a continuación y que se refieren: al carácter defectuoso del producto (art. 10.2); al nexo causal (art. 10.3); y al carácter defectuoso y nexo causal en determinadas circunstancias que habrán de valorarse por un órgano jurisdiccional (art. 10.4).

El art. 10.2 señala que se presumirá el carácter defectuoso del producto cuando se dé alguna de las siguientes circunstancias:

- no se hayan exhibido las pruebas requeridas por la autoridad judicial
- se haya demostrado que el producto no cumple los requisitos obligatorios de seguridad
- se haya demostrado que el daño fue causado por un mal funcionamiento evidente del producto

De estas tres causas lo cierto es que la última supone de alguna manera probar el propio carácter defectuoso del producto; también la prueba de la falta de cumplimiento de los requisitos obligatorios se acerca bastante a la prueba del propio defecto. Solo la primera puede considerarse como una auténtica presunción.

En cuanto a cuáles son los requisitos obligatorios de seguridad, parece que ha de entenderse por tales los establecidos en la normativa de seguridad del producto. En general, ha de recordarse que el sistema de nuevo enfoque se limita a establecer requisitos de seguridad generales y que el cumplimiento de las normas técnicas o códigos de buenas prácticas en ocasiones supone la presunción de que se cumplen tales requisitos generales. En cualquier caso, bastaría con demostrar el incumplimiento de un requisito obligatorio, por ejemplo, la falta de elaboración de la documentación técnica. Ahora bien, ha de tenerse en cuenta que también es necesaria la prueba del nexo causal y parece que esta prueba ha de vincularse con el correspondiente incumplimiento. Es decir, habrá de demostrarse que fue precisamente la falta de elaboración de la documentación técnica la que causó el defecto del producto.

Respecto de la presunción relativa al nexo causal, el apartado 3 del art. 10 de la Directiva señala que se presumirá la causalidad cuando se compruebe que el producto es defectuoso y que el daño es compatible con tal defecto. Se trata también de una presunción que puede calificarse de débil, en cuanto requiere la prueba del defecto, del daño y la compatibilidad del daño con el defecto. No se requiere, por tanto, una prueba inamovible del nexo causal, pero si una cierta razonabilidad.

El sistema de presunciones se cierra con la posibilidad de que, en caso de que haya dificultades excesivas, debido a la complejidad técnica o científica para demostrar bien el carácter defectuoso, bien el nexo causal, y siempre que el demandante aporte pruebas de que el producto contribuyó a los daños y de que es probable que el producto sea defectuoso o que su carácter defectuoso sea una causa probable de los daños, el órgano

jurisdiccional nacional pueda presumir el carácter defectuoso o el nexo causal o ambas cosas (art. 10.4).

En todo caso, el demandado puede refutar cualquiera de las presunciones (art. 10.5).

### 4.3.4. Sujetos responsables

La Directiva contempla como posibles sujetos responsables a los operadores económicos entre los que se incluye el fabricante, el representante autorizado, el importador, al prestador de servicios de tramitación de pedidos a distancia y al distribuidor.

La responsabilidad principal, de acuerdo con el art. 8 de la Directiva incumbe al fabricante. En el concepto de fabricante se incluye también el de un componente siempre que el daño sea causado por tal componente. Asimismo, también se considerará fabricante a la persona que modifique el producto, siempre que la modificación sea considerada sustancial de acuerdo con las normas de la Unión y que el daño no esté relacionado con alguna parte no afectada por la modificación (art. 11.1g). En el caso de los sistemas de IA, el apartado 23 del art. 3 RIA considera modificación sustancial: "un cambio en un sistema de IA tras su introducción en el mercado o puesta en servicio que no haya sido previsto o proyectado en la evaluación de la conformidad inicial realizada por el proveedor y a consecuencia del cual se vea afectado el cumplimiento por parte del sistema de IA de los requisitos establecidos en el capítulo II, sección 2, o que dé lugar a una modificación de la finalidad prevista para la que se haya evaluado el sistema de IA en cuestión".

En caso de que el fabricante esté fuera de la Unión responderá el importador del producto, esto es, la persona que introduce el producto en la UE o el representante autorizado del fabricante, que es la persona que recibe un mandato del fabricante para actuar en su nombre.

Cuando ninguno de los anteriores esté establecido en la Unión responderá el prestador de servicios de tramitación de pedidos a distancia.

Por último, cuando no pueda identificarse a ninguno de los anteriores responderá el distribuidor, salvo que consiga identificar al operador económico o a la persona que le suministró al producto.

El sistema se completa con una serie de reglas en el caso de que haya varios responsables. En general, el art 12 señala que cuando dos o más operadores sean responsables de los mismos daños responderán conjunta y solidariamente. Esto podrá ocurrir, por ejemplo, cuando el daño se debe a un defecto del producto atribuible, tanto al fabricante del producto, como al fabricante de un componente. O cuando el defecto sea atribuible al fabricante y al empresario, que realiza una modificación sustancial. Ahora bien, el art. 11 contempla una posibilidad de eximir de responsabilidad al fabricante del programa informático defectuoso.

Por otro lado, el art. 13 contempla que la responsabilidad del operador se puede reducir o incluso anular cuando el defecto del producto se deba a la culpa de la persona perjudicada o de una persona de la que deba responder.

En cambio, señala que la responsabilidad no se reducirá cuando el daño se deba tanto al carácter defectuoso del producto como al acto u omisión de un tercero.

## 5. Responsabilidad civil por daños derivados de la IA en el ámbito laboral

### *5.1. Articulación de los diversos bloques normativos en materia de seguridad y de responsabilidad*

De cuanto se ha visto hasta el momento resulta que, tanto en el bloque normativo relativo a la seguridad en el uso de estos sistemas de IA, como en el bloque normativo referido a la responsabilidad, existen diversas normas que convergen.

Se ha visto cómo el RIA impone una serie de obligaciones dirigidas fundamentalmente a los sistemas calificados como de alto riesgo, que lógicamente han de ser observadas por el proveedor de estos sistemas. Sin embargo, la observación de esta normativa no desplaza el necesario cumplimiento de la normativa en materia de prevención de riesgos laborales, la cual, impone a los fabricantes de todo tipo de productos la obligación de garantizar la seguridad de sus equipos. Junto a ello, ha de tenerse en cuenta que el propio empresario se erige en el principal sujeto obligado frente al fabricante.

Esta cierta heterogeneidad en la definición de las obligaciones, motiva que en caso de que se cause un daño se abra una pluralidad de vías de reclamación. En efecto, según se ha visto, la UE ha elaborado una normativa específica dirigida a facilitar el resarcimiento de tales daños. Esta normativa tiene en cuenta las obligaciones específicas impuestas por el RIA a la hora de calificar un producto como defectuoso. Sin embargo, esta normativa no desplaza la aplicación del régimen general de responsabilidad civil previsto por los legisladores nacionales, en particular, no impide que se reclame en virtud del régimen de responsabilidad contractual o extracontractual.

A la luz de todo lo anterior, cuando un trabajador sufra un daño como consecuencia del uso de un sistema de IA se abre un abanico complejo de posibilidades de reclamación. Como se verá a continuación, se puede plantear la reclamación del trabajador frente al empresario y proveedor. Asimismo, puede ocurrir que sea el empresario el que reclame frente al proveedor del sistema de IA. Todo ello, con base en diversos regímenes de responsabilidad, así, cabría plantearse la posibilidad de reclamar con base en la Directiva de productos defectuosos o con base en el régimen común de responsabilidad civil contractual y extracontractual.

### *5.2. Reclamación del trabajador por daños*

En caso de que el trabajador sufriera daños derivados del uso de la IA en el ámbito laboral podría articular la correspondiente demanda de responsabilidad civil para reclamar tales daños. Esta demanda se puede dirigir frente al proveedor del sistema de IA o frente

al empresario o frente a ambos. En la práctica, la exigencia con que se formula la obligación de seguridad del empresario hace prever que lo más habitual sea que se reclame frente al empresario y que este acabe asumiendo la responsabilidad por los daños derivados del sistema de IA.

De acuerdo con lo visto hasta aquí, se detalla a continuación el abanico de vías de reclamación por las que puede optar.

### 5.2.1. Reclamación en virtud del régimen específico de la Directiva de productos defectuosos

En primer lugar, el trabajador que sufre un daño como consecuencia de la utilización de un sistema de IA puede dirigir su demanda frente al fabricante del producto que ha causado el daño. Para articular tal reclamación podría acudir al régimen de responsabilidad por productos defectuosos que se acaba de explicar. En este caso, el trabajador podría reclamar los daños tasados en el art. 6. Ahora bien, la aplicación de este régimen no impediría que el trabajador pudiera reclamar frente al fabricante por daños no cubiertos conforme al régimen específico.

Por otro lado, será muy habitual en el ámbito laboral que exista una pluralidad de sujetos responsables. Según se ha visto el proveedor, será generalmente el responsable del daño, si bien, puede ocurrir que existan otros fabricantes de determinados componentes o que incluso el propio empresario pueda convertirse en fabricante y tener que responder conforme a la Directiva de responsabilidad por productos defectuosos.

Así, puede que el daño se haya debido a un componente del producto; esto es, a un archivo o servicio conexo vinculado, por ejemplo, a una máquina que no ha funcionado correctamente. En este caso responderá el fabricante del componente y junto a él también puede responder el fabricante del producto sisempre que estuviera bajo su control.

También puede ocurrir que el daño se haya debido a la falta de actualizaciones o mejoras. En este caso el responsable será el fabricante salvo que el producto no esté bajo su control.

En el caso de que el daño se haya debido a la actuación del empresario, que ha modificado sustancialmente el producto, responderá el empresario, como si fuese el propio fabricante, salvo que el daño se deba a una parte del producto no afectada por la modificación.

Cuando varios de estos sujetos sean responsables responderán conjuntamente.

Por último, cabe la posibilidad de que el daño se deba al defecto del producto y la culpa de la persona perjudicada, en nuestro caso, el trabajador. En este supuesto el art. 13.2 prevé que se pueda reducir o anular la responsabilidad del fabricante. En cualquier caso, será necesaria la culpa, no bastaría un mero acto no culpable, puesto que el art. 13 en su apartado1 lo señala expresamente.

#### 5.2.2. Reclamación en virtud del régimen de responsabilidad civil extracontractual frente al fabricante: la prueba de la negligencia y el proyectado régimen de presunciones

La Directiva de productos defectuosos no impide que el trabajador pueda articular una acción de responsabilidad civil ex art. 1902 Cc. En virtud de este régimen, el perjudicado podrá reclamar todos los daños sufridos, sin limitación en cuanto a su tipología. Para ello, el trabajador deberá demostrar la negligencia del fabricante, lo cual, en ocasiones puede resultar ciertamente difícil para un particular.

Tal y como se ha señalado la propuesta de Reglamento de Responsabilidad por sistemas de IA, pretende aligerar la carga de la prueba y contiene una serie de presunciones para facilitar la prueba del incumplimiento y del nexo causal centrado fundamentalmente en los sistemas de IA de alto riesgo y en coherencia con el régimen de obligaciones impuesto a los proveedores de estos sistemas por el Reglamento de IA.

### *5.3. Reclamación del trabajador frente al empresario*

Tal y como se ha señalado, las obligaciones de seguridad del fabricante de productos, incluyan o no sistemas de IA, no desplazan el resto de obligaciones que incumben al empresario. De tal manera que el empresario sigue siendo el máximo garante de la seguridad y salud de los trabajadores y deberá asumir los daños derivados del incumplimiento de su obligación de seguridad.

Además, habrá de tenerse en cuenta que el empresario que integra sistemas de IA en su organización productiva asume una serie de obligaciones como responsable del despliegue que, en caso de incumplimiento que dé lugar a daños, también podrán justificar la correspondiente reclamación de responsabilidad civil.

También, por último, el empresario puede asumir el rol de fabricante por haber introducido modificaciones sustanciales en el producto, lo cual puede permitir al trabajador que sufre un daño como consecuencia del uso de un producto, dirigir también su reclamación de responsabilidad civil por los daños sufridos frente al empresario.

En la práctica, la exigencia con que es definida la obligación de seguridad del empresario y las facilidades procesales en materia de prueba, determinan que sea habitual que el trabajador reclame los daños sufridos como consecuencia de un accidente del empresario y que no dirija su reclamación frente a terceros, como el fabricante o proveedor, ajenos a la relación laboral.

Desde un punto de vista teórico el empresario cabría plantearse si el empresario puede alegar la excepción de falta de litisconsorcio y traer al proceso laboral al fabricante. En este caso, conforme con lo previsto en el art. 1 LRJS el orden de lo social es el competente para conocer de todas las acciones, también frente a terceros ajenos a la relación laboral, relacionadas con daños derivados de accidente. El competente para examinar la posible responsabilidad del fabricante del producto o componente, incluso aplicando el régimen específico de la directiva de productos defectuosos sería el juez de lo social sería

competente. Con todo, no es pacífico que el empresario pueda alegar la mencionada excepción, ya que no se admite en los casos de solidaridad impropia. En efecto, en el ámbito de la responsabilidad por accidentes de trabajo, suele ser habitual que haya una pluralidad de sujetos responsables que deban asumir las consecuencias dañosas del accidente; si bien, la jurisprudencia del TS ha entendido que cuando la responsabilidad de tales sujetos no deriva de una norma o de un pacto expreso se está ante una solidaridad impropia; se trata de supuestos en los que existe una pluralidad de sujetos que han concurrido en la causación del daño y respecto de los cuales no es posible individualizar sus responsabilidades[24]. Este tipo de responsabilidad solidaria tiene efectos procesales, por cuanto la reclamación frente a alguno de los potenciales sujetos responsables no interrumpe la prescripción respecto del resto y tampoco permite la alegación de la excepción de litisconsorcio pasivo necesario; de manera que cabría interponer la demanda solo frente a uno de los sujetos responsables, en nuestro caso generalmente el empresario, sin margen para que este alegue la excepción de falta de litisconsorcio.

Únicamente cabría plantearse un régimen de responsabilidad solidaria propia, cuando la responsabilidad del empresario se articule por la vía de la Directiva de productos defectuosos, que, según se ha visto, puede llevar a que el empresario sea considerado como un operador económico. En este caso la Directiva impone una responsabilidad de carácter solidaria y añade que el derecho de repetición se articulará conforme a la legislación nacional. Sin embargo, en la práctica no resultará muy habitual que se sustente la responsabilidad del empresario en este particular régimen, sino que lo más frecuente será basarlo en el incumplimiento de su obligación contractual de seguridad.

### *5.4. Reclamación del empresario frente al proveedor*

Puede, por último, ocurrir que sea el empresario el que plantee una reclamación de responsabilidad civil frente al proveedor.

Así sucederá cuando el empresario sea la persona perjudicada conforme a la Directiva de productos defectuosos. Recuérdese que esta Directiva le permite reclamar solo por los daños personales, incluidos los morales. No podrá, sin embargo, reclamar en virtud de este régimen los daños sufridos en el propio producto o en otros productos destinados al ámbito profesional. Ahora bien, el empresario podrá reclamar los daños sufridos en las instalaciones o equipos de trabajo por el mal funcionamiento de un producto con arreglo a otros regímenes de responsabilidad contractual o extracontractual.

[24] STS de 6 de mayo de 2021 (rec. 2611/2018), comentada en: CASAS BAAMONDE, M.E.: "Responsabilidad empresarial por accidentes de trabajo en contratas y subcontratas. La responsabilidad civil solidaria como solidaridad impropia", *Revista de Jurisprudencia Laboral,* núm. 6, 2021, disponible en: https://www.boe.es/biblioteca_juridica/anuarios_derecho/articulo.php?id=ANU-L-2021-00000001310. Seguida por SSTJS (Castilla y León), de 1 de octubre de 2021 (rec. 882/2021) y (Canarias), de 2 de noviembre de 2021, (rec. 115/2021).

En el caso de que el empresario asuma la totalidad de la reparación, debido a que el trabajador haya dirigido la demanda solo frente a él y no se haya podido incorporar al fabricante al proceso, surge la duda de cómo podrá reclamar del fabricante la indemnización por él asumida. En principio, si la responsabilidad del empresario tiene su origen en la Directiva de productos defectuosos, la propia Directiva señala que la regla es la solidaridad y, en consecuencia, si el empresario ha asumido el total de la indemnización deberá dirigir frente al fabricante la correspondiente acción de repetición

Mayores dudas se suscitan en el caso de que el empresario haya resultado responsable no conforme a la Directiva, sino con base en su obligación general de seguridad. En este supuesto se plantea si, en caso de que el fabricante también haya contribuido al daño, el empresario puede reclamar parte de la indemnización ya satisfecha y conforme a qué régimen. En principio, la Directiva de productos defectuosos solo permite reclamar los daños personales, de tal modo que, salvo que haya existido condena conjunta del empresario y del fabricante conforme con este régimen. En otro caso, el empresario cabría plantearse la posibilidad de que el empresario reclame del proveedor en virtud del régimen de responsabilidad contractual, en caso de que medie contrato entre ambos.

## 6. Bibliografía

ARTIÑANO MARRA, P. y SÁNCHEZ ORO, J.: "Responsabilidad por el funcionamiento de sistemas de inteligencia artificial: los desafíos de la "documentación técnica del Reglamento de Inteligencia Artificial", *Derecho Digital e Innovación,* núm. 20, 2024.

ATAZ LÓPEZ, J.: "Daños causados por las cosas una nueva visión a raíz de la robótica y de la inteligencia artificial", *Working Paper*, 4/2020, UB, disponible en: http://hdl.handle.net/2445/169850;

ATIENZA NAVARRO, M.L.: "La responsabilidad civil por daños causados por inteligencia artificial. Estado de la cuestión" en AA.VV.: *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* Aranzadi, Pamplona, 2024.

– *Daños causados por inteligencia artificial y responsabilidad civil,* Atelier, Barcelona, 2022;

CASAS BAAMONDE, M.E.: "Responsabilidad empresarial por accidentes de trabajo en contratas y subcontratas. La responsabilidad civil solidaria como solidaridad impropia", *Revista de Jurisprudencia Laboral,* núm. 6, 2021, disponible en: https://www.boe.es/biblioteca_juridica/anuarios_derecho/articulo.php?id=ANUL-2021-00000001310. Seguida por SSTJS (Castilla y León), de 1 de octubre de 2021 (rec. 882/2021) y (Canarias), de 2 de noviembre de 2021, (rec. 115/2021).

DE SILVA LÓPEZ DE LETONA, J.: "Responsabilidad por daños causados por sistemas de inteligencia artificial", *Derecho Digital e Innovación,* núm. 11, 2022.

DIRECCIÓN GENERAL DE REDES DE COMUNICACIÓN, CONTENIDO Y TECNOLOGÍAS (COMISIÓN EUROPEA): *Directrices éticas para una IA*

*fiable,* 2019, Disponible en: https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1.

EBERS, M.: "La utilización de los agentes electrónicos inteligentes en el tráfico jurídico. ¿Necesitamos reglas especiales en el Derecho de la responsabilidad civil?", *Indret,* núm. 3, 2016.

EGUSQUIZA BALMASEDA, M.E.: "Marco normativo general y propuestas de regulación en la responsabilidad civil" en AA.VV. (dir.: RODRÍGUEZ SANZ DE GALDEANO, B. y EGUSQUIZA BALMASEDA, M.E.): *Inteligencia Artificial y prevención de riesgos laborales: obligaciones y responsabilidades,* Tirant lo Blanch, Valencia, 2023.

ESPÍN ALBA, I.: "Sesgos discriminatorios en la toma automatizada de decisiones en la contratación y protección de datos" en AA.VV.: *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* Aranzadi, Pamplona, 2024.

ESTEVE PARDO, J.: *Técnica, riesgo y Derecho. Tratamiento del riesgo tecnológico en el Derecho ambiental,* Ariel Derecho, 1999. Del mismo autor: "La regulación d ellos riesgos: gestionar la incertidumbre", *El Cronista del Estado Social y Democrático de Derecho,* núm. 96-97 (octubre-noviembre), 2021.

EVANGELIO LLORCA, R.: "Causalidad y responsabilidad civil por daños ocasionados por sistemas de inteligencia artificial: las presunciones de causalidad en las propuestas normativas de la UE", en AA.VV.: *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* Aranzadi, Pamplona, 2024.

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

GOÑI SEIN, J.L.: "El Reglamento UE de inteligencia artificial y su interrelación con la normativa de seguridad y salud en el trabajo", en AA.VV.: (dir.: EGUSQUIZA BALMASEDA, M.E. y RODRÍGUEZ SANZ DE GALDEANO, B.): *Inteligencia Artificial y Prevención de Riesgos Laborales: Obligaciones y Responsabilidades",* Tirant lo Blanch, Valencia, 2023.

GOÑI SEIN, J.L.: "Sistemas de inteligencia artificial y prevención de los riesgos laborales: obligaciones del proveedor y del empresario", *Labos*, (en este nº monográfico).

JORQUI AZOFRA, M.: "El concepto legal de producto a la luz de la nueva propuesta de directiva sobre responsabilidad por los daños causados pro productos defectuosos", en AA.VV. (dir.: RODRÍGUEZ SANZ DE GALDEANO, B. y EGUSQUIZA BALMASEDA, M.E.): *Inteligencia Artificial y prevención de riesgos laborales: obligaciones y responsabilidades,* Tirant lo Blanch, Valencia, 2023.

PARRA LUCÁN, M.A.: "Responsabilidad civil por productos defectuosos" en REGLERO CAMPOS, L.F. (COORD.): *Tratado de Responsabilidad civil,* Aranzadi, Pamplona, 2014.

JORQUI AZOFRA, M.: *Responsabilidad por daños causados por productos y sistemas de inteligencia artificial,* Dykinson, Madrid, 2023.

LLANO ALONSO, F.H.: "Ética(s) de la inteligencia artificial y derecho consideraciones a propósito de los límites y la contención del desarrollo tecnológico. *Derechos y libertades,* núm. 51, época II, junio 2024, pp. 177 y ss, disponible en https://e-revistas.uc3m.es/index.php/DYL/article/view/8587/6595.

LLORENS ESPADA, J.: "La inteligencia artificial para la mejora de la seguridad y salud laboral y su encaje en el marco regulatorio europeo"*, Trabajo y Derecho,* núm. 19, 2024.

MERCADER UGUINA, J.: "El Reglamento de Inteligencia Artificial: frecuentemos el futuro", *Colección Briefs,* 20/3/2024, AEDTSS, disponible en: https://www.aedtss.com/el-reglamento-de-inteligencia-artificial-frecuentemos-el-futuro/, que el nuevo Reglamento se basa en el triángulo de oro: aproximación desde el riesgo, garantías y responsabilidades.

NAVAS NAVARRO, S.: *Daños ocasionados por sistemas de inteligencia artificial. Especial atención a su futura regulación,* Comares, Granada, 2022

– "Régimen europeo en ciernes en materia de responsabilidad deriva de los sistemas de Inteligencia artificial", *Revista CESCO de Derecho del Consumo,* núm. 44, 2022, pp. 43 y ss.

PEÑA LÓPEZ, F.: "Responsabilidad objetiva y subjetiva en las propuestas legislativas europeas sobre responsabilidad civil aplicables a la inteligencia artificial" en AA.VV.: *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* Aranzadi, Pamplona.

RODRÍGUEZ SANZ DE GALDEANO, B.: "La responsabilidad empresarial por accidentes vinculados a la inteligencia artificial", *Trabajo y Derecho: nueva revista de actualidad y relaciones laborales,* núm. extra 19, 2024.

– "Los sistemas de inteligencia artificial en el ámbito laboral y el marco regulador europeo de seguridad del producto" (dir.: EGUSQUIZA BALMASEDA, M.E. y RODRÍGUEZ SANZ DE GALDEANO, B.): *Inteligencia Artificial y Prevención de Riesgos Laborales: Obligaciones y Responsabilidades",* Tirant lo Blanch, Valencia, 2023.

– "Obligaciones del empresario en materia de prevención de riesgos laborales derivadas de la utilización de sistemas de IA", *Revista Galega de Dereito Social,* núm. 18, 2023.

*Las responsabilidades de los fabricantes en materia de prevención de riesgos laborales,* Lex Nova, Valladolid, 2005.

RUBÍ PUIG, A.: "Inteligencia artificial y daños indemnizables", en AA.VV.: AA.VV.: *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* Aranzadi, Pamplona, 2024.

VAZQUEZ DE CASTRO, E.: "Aproximación a la responsabilidad derivada de los riesgos de la inteligencia artificial en Europea", AA.VV. (SOLAR CAYÓN, J.I.): *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho,* Editorial Universidad de Alcalá, 2020.

# Sistema de responsabilidades por el uso de la inteligenia 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".

# Reglamento de inteligencia artificial e intervención pública en las relaciones laborales[*]

## Artificial intelligence act and public intervention in labor relations

José María Goerlich Peset
*Catedrático de Derecho del Trabajo y Seguridad Social*
*Universitat de València*
ORCID ID: 0000-0002-2910-2153

doi: 10.20318/labos.2024.9040

*Resumen:* Aunque la atención de los juristas del trabajo, y del propio legislador, se ha centrado en el uso de la inteligencia artificial por las empresas, en el marco del contrato de trabajo, lo cierto es que en las dos últimas décadas hemos asistido a un imparable incremento de su utilización por las entidades públicas, administrativas y judiciales. El presente trabajo describe, en primer lugar, la evolución normativa interna en relación con esta última. Se analiza, en segundo lugar, el impacto que sobre la situación actual puede tener la aprobación del Reglamento de Inteligencia Artificial. Finalmente, se valora la situación resultante. En este último terreno, se destacan las insuficiencias existentes, sobre todo en relación con el uso administrativo de la IA, y se ofrecen algunas posibilidades para superarlas.

*Palabras clave:* Inteligencia artificial, relaciones laborales, administración laboral, proceso laboral.

*Abstract:* Although the attention of labor jurists, and of the legislator himself, has been focused on the use of artificial intelligence by companies, within the framework of the employment contract, the fact is that in the last two decades we have witnessed an unstoppable increase in its use by public, administrative and judicial entities. This paper describes, firstly, the internal normative evolution in relation to the latter. Secondly, it analyzes the impact that the approval of the Artificial Intelligence Act may have on the current situation. Finally, the resulting situation is assessed. In this last area, the existing inadequacies are highlighted, especially in relation to the administrative use of AI, and some possibilities are offered to overcome them.

*Keywords:* Artificial intelligence, labor relations, labor administration, labor process.

[*]Trabajo realizado en el marco del proyecto de investigación, Algoritmos extractivos y neuroderechos. Retos regulatorios de la digitalización del trabajo (PID2022-139967NB-I00).

El título original que tenía esta ponencia en el programa del Congreso era "Las sanciones automatizadas en la Inspección de Trabajo". En mi intervención oral, ya amplié la óptica con el ánimo de incluir el conjunto en las decisiones administrativas automatizadas que son, como sabemos, muy frecuentes en el orden social, y analizar cómo impacta en su régimen jurídico el nuevo Reglamento de Inteligencia Artificial (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, DOUE 12 julio 2024 –en adelante, RIA–). En esta versión escrita, que, por otra parte, tiene un tono similar al de la presentación oral, añadiré algunas consideraciones adicionales en relación con las posibilidades de utilización de la inteligencia artificial en el marco del proceso laboral.

Mi plan es abordar el tema en tres partes. La primera describe brevemente la situación normativa interna en el momento de la aprobación del Reglamento europeo. La segunda analiza sus disposiciones para determinar cómo podrían afectar a aquella. Finalmente, intento valorar si el conjunto resultante es o no adecuado en atención a los intereses que están en juego.

## 1. La creciente presencia de la digitalización en los entes públicos

Diría que, en los últimos 30 años, hemos asistido a una creciente incorporación de las nuevas tecnologías, no solo en las empresas a las que se ha dedicado la atención de las ponencias anteriores, sino también en las entidades públicas.

Por lo que se refiere a las Administraciones Públicas en general, podemos fijar convencionalmente el arranque de este proceso en la reforma del procedimiento administrativo de 1992. Según su Exposición de Motivos, la Ley 30/1992, de 26 de noviembre, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común, estaba llamada a protagonizar la "abierta incorporación de las técnicas informáticas y telemáticas en la relación ciudadano-Administración". Su art. 45, dedicado a la "incorporación de medios técnicos", afirmaba que las Administraciones Públicas impulsarían "el empleo y aplicación de las técnicas y medios electrónicos, informáticos y telemáticos, para el desarrollo de su actividad y el ejercicio de sus competencias, con las limitaciones que a la utilización de estos medios establecen la Constitución y las Leyes" (apartado 1); y a continuación preveía que "los programas y aplicaciones electrónicos, informáticos y telemáticos que vayan a ser utilizados por las Administraciones Públicas para el ejercicio de sus potestades, (habrían) de ser previamente aprobados por el órgano competente, quien deberá difundir públicamente sus características".

Las posteriores reformas siguieron avanzando en este sentido. La regulación de la Ley 30/1992 fue sustituida por la contenida en el art. 39 Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos, que preveía ya la posibilidad de adoptar decisiones automatizadas en el ámbito administrativo. No se trataba solo de la utilización de estos medios sino de que fueran estos los que adoptaran la decisión. Para este caso, se preveía que "deberá establecerse previamente el órgano u órganos

competentes, según los casos, para la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, auditoría del sistema de información y de su código fuente. Asimismo, se indicará el órgano que debe ser considerado responsable a efectos de impugnación". Esta regulación se ha incorporado al art. 41.2 de la vigente Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público. En su apartado primero, el precepto define, además, lo que debe entenderse por actuación administrativa automatizada, que resulta ser "cualquier acto o actuación realizada íntegramente a través de medios electrónicos por una Administración Pública en el marco de un procedimiento administrativo y en la que no haya intervenido de forma directa un empleado público". Existen algunas concreciones adicionales de este régimen, tampoco muchas, en el art. 13 del Reglamento de actuación y funcionamiento del sector público por medios electrónicos, aprobado por RD 203/2021, de 30 de marzo.

Es posible advertir cómo en este proceso se ha ido redimensionando el régimen jurídico. Y ello, en un doble sentido. De un lado, en relación con las actuaciones afectadas, se ha pasado de la genérica utilización de "programas y aplicaciones electrónicos, informáticos y telemáticos… por las Administraciones Públicas para el ejercicio de sus potestades" (Ley 30/1992) a la existencia de una actuación "realizada íntegramente a través de medios electrónicos" sin intervención directa de humano (Ley 40/2015). De otro, por lo que se refiere a las obligaciones informativas a cargo de las Administraciones, si hasta 2007 había que comunicar los programas y aplicaciones electrónicos, informáticos y telemáticos que fueran a ser utilizados por las Administraciones Públicas para el ejercicio de sus potestades, desde entonces la información parece más limitada. Volveré al final sobre este tema.

Las autoridades que ejercitan competencias en el ámbito social no han sido ajenas a esta evolución, como he analizado con más detalle en otro lugar (Goerlich Peset, 2021). Con el soporte que les proporcionaban las citadas normas han ido modernizando sus estructuras y sumándose a la imparable digitalización. Las actuaciones automatizadas se han ido incorporando, primero, a la gestión de la Seguridad Social y, luego, a la Inspección de Trabajo y Seguridad Social. En el primer terreno, desde mediados de la década de los 90, informática y telemática se han puesto al servicio de las actividades instrumentales que desarrolla la Seguridad Social: actos de encuadramiento y recaudación han sido afectados por sucesivas normas que implican su completo tránsito a la esfera digital. Esto proceso se corona a mediados de la siguiente década, cuando el sistema tradicional de recaudación basado en la autoliquidación se sustituye por el actualmente vigente: la aprobación de la Ley 34/2014, de 26 de diciembre, de medidas en materia de liquidación e ingreso de cuotas de la Seguridad Social, presupone, de un lado, una formidable capacidad de gestión de la información por parte de las Entidades Gestoras y garantiza, de otro, un igualmente formidable caudal de datos a disposición del uso administrativo de la inteligencia artificial.

La incorporación normativa de la posibilidad de adoptar decisiones automatizadas se produce en 2009, para las prestaciones de desempleo (disp. adic. 46ª LGSS-1994, redacción aprobada por RDL 10/2009, de 13 de agosto) y se amplía después, por la

ya citada Ley 30/2014, al conjunto de las prestaciones, excluidas las pensiones en su modalidad no contributiva. Incorporadas estas reglas al art. 130 del texto refundido aprobado en 2015, este ha vuelto a ser objeto de modificación por el RDL 2/2021, de 26 de enero, de refuerzo y consolidación de medidas sociales en defensa del empleo. En su nueva redacción, actualmente vigente, extiende la posibilidad de adoptar decisiones automatizadas al conjunto de la gestión del sistema de Seguridad Social, con la única salvedad de las prestaciones no contributivas. Por lo demás, aunque separado del régimen administrativo común por razones vinculadas a la tradición, las previsiones de su segundo párrafo para las decisiones automatizadas en el ámbito de la Seguridad Social reproducen prácticamente al pie de la letra las previsiones del art. 45 Ley 40/2015.

Algo parecido ha venido ocurriendo en el ámbito de la Inspección de Trabajo. Aunque aparentemente tiene poco que ver con el tema, llama mucho la atención la propia evolución de la definición legal del sistema de Inspección en la ley reguladora de 1997 y en la que la sustituyó en 2015. En la primera, Ley 42/1997, de 14 de noviembre, Ordenadora de la Inspección de Trabajo y Seguridad Social, el art. 1.1 nos decía que venía constituido por "el conjunto de principios legales, normas, órganos, funcionarios y medios materiales que contribuyen al adecuado cumplimiento de las normas" que entran en su ámbito de competencia. Veinte años después, el art. 1.1 Ley 23/2015, de 21 de julio, Ordenadora del Sistema de Inspección de Trabajo y Seguridad Social, indica que lo está por "el conjunto de principios legales, normas, órganos, personal y medios materiales, incluidos los informáticos, que contribuyen al adecuado cumplimiento de la misión que tiene encomendada". Encontramos una diferencia significativa entre los dos preceptos: la relevante referencia a "los medios informáticos" que no eran aludidos en la norma de 1997 y aparecen en la de 2015. Se resalta de este modo su papel central en la actuación inspectora. Por supuesto, no puede dejar de traerse a colación la reforma de la LISOS por el ya citado RDL 2/2021. En su redacción vigente, el 53.1.a) LISOS prevé, nada más y nada menos, la posibilidad de que un acta de inspección sea emitida a través de una decisión administrativa automatizada generada por vía algorítmica. Con posterioridad, el RD 688/2021, de 3 de agosto, ha modificado el Reglamento general sobre procedimientos para la imposición de sanciones por infracciones de orden social y para los expedientes liquidatorios de cuotas de la Seguridad Social (RD 928/1998) y establecido una serie de reglas específicas para estas actas automatizadas –que, por cierto, hasta donde yo alcanzo, aún no se han puesto en marcha–.

Me he centrado, hasta ahora, en los usos administrativos de la IA que han cristalizado en concretas normas jurídicas, legales o reglamentarias. Pero las entidades que hemos considerado hacen uso también de la IA en terrenos diferentes que no dejan trazas formales –o las dejan de otra naturaleza–. Este uso no da lugar a decisiones administrativas automatizadas en el sentido expuesto sino que se mueve en el terreno de la inteligencia artificial predictiva que viene siendo utilizada, tanto por la Seguridad Social como por la Inspección de Trabajo, para la detección de posibles espacios de «oscuridad». Sabemos, en este sentido, que la Seguridad Social emplea este tipo de IA para detectar supuestos o bolsas de fraude, en el terreno de la recaudación y también en el del control de prestacio-

nes –en relación, por ejemplo, con la incapacidad temporal–. Sin embargo, las noticias que tenemos de este uso son menores. En los casos de decisiones automatizadas formales la sede virtual (https://sede.seg-social.gob.es/wps/portal/sede/sede/Inicio/NormativayLegislacion) incluye un catálogo de resoluciones administrativas que informan sobre la utilización de determinadas herramientas algorítmicas y cuáles son sus características generales. Por el contrario, en el caso de la IA predictiva, la información que tenemos es indirecta a través de la literatura especializada (Aibar Bernard, 2020; Redondo Rincón, 2020). Es seguro, sin embargo, que se está generalizando como ha puesto de manifiesto la AISS en fecha relativamente reciente (2019, p. 34).

En cuanto a su uso en el ámbito de la Inspección, sabemos que existe una "herramienta de lucha contra el fraude", o sea un conjunto de aplicaciones que permiten hacer minería de datos y, tras su tratamiento, detectar sectores o espacios en los que concentrar la actuación inspectora. Tenemos bastante información sobre sus utilidades prácticas. A través del buscador del BOE, detectamos su presencia en los sucesivos planes que la Inspección de Trabajo ha ido publicando en los últimos años. Se alude a ella, por ejemplo, en el Plan estratégico 2018/2020 (BOE 19 abril 2018) y en el inmediato Plan director por un trabajo digno 2018-2020 (BOE 28 julio 2108). El último Plan estratégico conocido, para el período 2021/2023 (BOE 3 diciembre 2021), contiene más de 30 referencias, tanto en los objetivos como en las actuaciones. Es más, tiene sustantividad orgánica como "unidad del OEITSS" (actuación 32.2), con funciones de planificación (actuaciones 30.1 y 30.2) y en el terreno de las políticas de personal (22.4) y las reformas orgánicas (23.3). Por lo que se refiere a los objetivos pueden destacarse los núms. 30 –"Mejorar la planificación de las actuaciones inspectoras, utilizando las herramientas más avanzadas de inteligencia artificial"– y 32 –"Adecuar las infraestructuras, medios materiales, los sistemas, redes, aplicaciones y equipamientos informáticos, así como fortalecer la Herramienta de Lucha contra el Fraude"–. En cuanto a las actuaciones, sale en varias de las recogidas en los diferentes Ejes –salarios (1.1); tiempo de trabajo (1.2); contratación (1.3), convenios (1.4), falsos autónomos (1.7); diferentes aspectos de PRL (2.1 y 2.3); trabajo no declarado (3.1); discriminación por razón de sexo (5.1) u otras causas (6.1); recaudación (8.1, 8.5); inaplicación (19.1)–.

Por último, también la actividad jurisdiccional se ha visto afectada por la digitalización (para un análisis de conjunto, Nores Torres, 2023 con detalladas referencias adicionales). La evolución de la redacción del art. 230 LOPJ es muy significativa del proceso. Si originalmente el precepto se refería al uso de "medios técnicos de documentación y reproducción, siempre que ofrezcan las debidas garantías de autenticidad", desde 1994 (LO 16/1994) se ha dado entrada al de "cualesquiera medios técnicos, electrónicos, informáticos y telemáticos, para el desarrollo de su actividad y ejercicio de sus funciones"; y, además, se ha posibilitado la relación entre interesados y órganos judiciales y garantizado la autenticidad de la documentación y la identificación de sus autores. Si inicialmente el uso de estos medios era voluntario, con posterioridad se ha hecho obligatorio para los órganos judiciales (LO 7/2015) y, después, para las personas que recurran a ellos (LO 4/2018).

Estos cambios en la norma orgánica han ido acompañados de otros más detallados a lo largo del período que se considera (entre otras disposiciones, leyes 18/2011 y 42/2015). Recientemente, el extensísimo RDL 6/2023, de 19 de diciembre, por el que se aprueban medidas urgentes para la ejecución del Plan de Recuperación, Transformación y Resiliencia en materia de servicio público de justicia, función pública, régimen local y mecenazgo, dedica la friolera de ¡100 artículos! a las diferentes cuestiones implicadas, estableciendo una regulación aplicable a todos los órdenes jurisdiccionales. Heredero en este punto del proyecto de ley de Medidas de Eficiencia Digital del Servicio Público de Justicia decaído como consecuencia de la disolución de las Cortes en primavera de 2023, esta parte del RDL 6/2023 se presenta como "una herramienta normativa completa, útil, transversal y con la capacidad suficiente para dotar a la Administración de Justicia de un marco legal, coherente y lógico en el que la relación digital se descubra como una relación ordinaria y habitual, siendo la tutela judicial efectiva en cualquier caso la prioridad absoluta, pero hallando bajo esta cobertura de normas y reglas un nuevo cauce, más veloz y eficaz, que coadyuvará a una mejor satisfacción de los derechos de la ciudadanía".

Entre otras cuestiones, la nueva regulación dedica su atención a la utilización de la inteligencia artificial en el marco de las actuaciones jurisdiccionales; y lo hace estableciendo importantes condicionantes (cfr. Nores Torres, 2024a, p. 232 ss. y 2024b, p. 101 ss.). De un lado, la posibilidad de actuaciones algorítmicas autónomas queda confinada a los aspectos instrumentales: conforme al art. 56.2 RDL 6/2023, únicamente se admite "la automatización de las actuaciones de trámite o resolutorias simples, que no requieren interpretación jurídica"; junto a ellas se admiten las actuaciones proactivas, esto es aquellas que se inician de forma autónoma "por los sistemas de información sin intervención humana, que aprovechan la información incorporada en un expediente o procedimiento de una Administración Pública con un fin determinado, para generar avisos o efectos directos a otros fines distintos, en el mismo o en otros expedientes, de la misma o de otra Administración Pública" (art. 56.3). Fuera de estos casos, el uso de la inteligencia artificial en el ámbito judicial queda limitado a las denominadas "actuaciones asistidas" (art. 57), que implican la posible generación de borradores de documentos complejos. Queda explícitamente prohibido que estos borradores constituyan, "sin validación de la autoridad competente", una resolución judicial o procesal; y se impone que solo puedan generarse "a voluntad del usuario" que además ha de poder modificarlos "libre y enteramente".

## 2. Reglamento de Inteligencia Artificial y actuaciones públicas

Descrito sucintamente el panorama de la situación previa, conviene abordar, como segunda cuestión, el impacto que puede desplegar sobre ella la reciente aprobación del Reglamento de Inteligencia Artificial. Avanzando la conclusión, diría que será limitado. Esto es así, creo, por dos razones: la primera es la propia noción de inteligencia artificial; la segunda, la forma que se aborda su regulación. Vayamos por partes.

Hay una cosa que está clara: el uso de la inteligencia artificial por las entidades públicas está afectada por el RIA. No puede alcanzarse otra conclusión si se miran las definiciones de proveedor de sistemas de IA [art.3.3) RIA] o responsable de su despliegue [art.3.4) RIA], que incluyen a "persona física, jurídica o autoridad pública, órgano u organismo". Es claro, por tanto, que subjetivamente las entidades públicas están afectadas, tanto si desarrollan este tipo de sistemas como si los utilizan "bajo su propia autoridad". Ello no obstante, es posible que no todos los usos que hemos considerado en el apartado anterior quepan en la noción de inteligencia artificial que utiliza el art. 3.1) RIA. Este precepto la conceptúa como "un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y 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 resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales". Entiendo que no existe equivalencia entre esta noción y la de decisión automatizada que hemos considerado en el apartado anterior.

Esta falta de equivalencia es doble. Por un lado, es posible que la noción jurídica de IA no incluya todos los supuestos de decisiones automatizadas concebibles, puesto que cabe pensar que existan algunas de estas, de mayor o menor sencillez, en las que no se detecta el requisito de la autonomía al que se refiere el precepto para considerar que estamos ante un sistema de IA. El extensísimo preámbulo del RIA permite alcanzar la conclusión de que los "distintos niveles de autonomía" y la "capacidad de adaptación tras el despliegue" excluyen de tal noción "los sistemas basados en las normas definidas únicamente por personas físicas para ejecutar automáticamente operaciones"; se hace preciso además que tengan "capacidad de inferencia" lo que "trasciende el tratamiento básico de datos al permitir el aprendizaje, el razonamiento o la modelización" (considerando 12). De este modo, muy probablemente no constituye un sistema de IA la herramienta informática que permite la emisión automatizada de una reclamación de deuda ante el impago de cotizaciones con base en el art. 33 LGSS. Por otro lado, pero ahora en un sentido diverso, la amplitud con la que se conciben los "resultados de salida" hace que los sistemas de IA desborden la estricta noción de decisiones automatizadas que ya hemos examinado. Las "predicciones, contenidos, recomendaciones o decisiones que puedan influir en entornos físicos o virtuales" a las que se refiere el art. 3.1) RIA permiten reconducir al ámbito de aplicación del nuevo reglamento actuaciones que no se toman en el marco de un proceso administrativo en el sentido del artículo 41 Ley 40/2015 –acaso las remisiones masivas relacionadas con vicisitudes de la contratación que hemos visto estos años–, así como la planificación de actividades inspectoras o de control por la seguridad social.

En otro orden de consideraciones, el modelo regulatorio del nuevo Reglamento se basa en una visión positiva de la IA. En esta línea, el preámbulo señala que "puede proporcionar ventajas competitivas esenciales a las empresas y facilitar la obtención de resultados positivos desde el punto de vista social y medioambiental en los ámbitos de la asistencia sanitaria, la agricultura, la seguridad alimentaria, la educación y la formación, los medios de comunicación, el deporte, la cultura, la gestión de infraestructuras, la ener-

gía, el transporte y la logística, los servicios públicos, la seguridad, la justicia, la eficiencia de los recursos y la energía, el seguimiento ambiental, la conservación y restauración de la biodiversidad y los ecosistemas, y la mitigación del cambio climático y la adaptación a él, entre otros, al mejorar la predicción, optimizar las operaciones y la asignación de los recursos, y personalizar las soluciones digitales que se encuentran a disposición de la población y las organizaciones" (§ 4). De este modo, como regla general, su distribución y uso son libres, salvo que se tope con una prohibición (art. 5 RIA) o, en razón de la materia a la que afecta, se esté en presencia de un sistema de alto riesgo, cuya puesta en funcionamiento queda sujeta a unas cautelas especiales (art. 6 RIA).

El repaso de estos condicionantes a distribución y despliegue de sistemas de IA muestra que las cuestiones que nos preocupan en este momento solo están presentes de manera limitada. Esto es bastante claro con las prácticas de IA prohibidas listadas en el art. 5 RIA, que no incluyen ninguna de las que ahora estamos analizando. En cuanto a la enumeración de sistemas de IA de alto riesgo contenida en el Anexo III, al que se remite el art. 6 RIA, sí encontramos algunas, aunque en modo alguno son coextensas con el ámbito que consideramos.

Por lo que se refiere al uso administrativo de la IA, las referencias del Anexo a la materia social no son muchas. Al margen la inclusión en el apartado 7 de los sistemas que afectan a "Migración, asilo y gestión del control fronterizo" –que pudieran tener cierto impacto colateral en las gestiones de la ITSS relacionadas con el trabajo de extranjeros–, cabe destacar la previsión de su apartado 5 que incluye entre los de alto riesgo los "sistemas de IA destinados a ser utilizados por las autoridades públicas o en su nombre para evaluar la admisibilidad de las personas físicas para beneficiarse de servicios y prestaciones esenciales de asistencia pública, incluidos los servicios de asistencia sanitaria, así como para conceder, reducir o retirar dichos servicios y prestaciones o reclamar su devolución". La razón de la clasificación de estos sistemas de IA entre los de alto riesgo la ofrece el § 58 de los considerandos del preámbulo del RIA. Su utilización en este terreno, se argumenta, "podría tener un efecto considerable en los medios de subsistencia de las personas y vulnerar sus derechos fundamentales, como el derecho a la protección social, a la no discriminación, a la dignidad humana o a la tutela judicial efectiva". Sin embargo, "el presente Reglamento no debe obstaculizar el desarrollo y el uso de enfoques innovadores en la Administración, que podrían beneficiarse de una mayor utilización de sistemas de IA conformes y seguros". La forma de equilibrar ambos intereses es su calificación como de alto riesgo.

Ello impacta, por supuesto, en la gestión de la seguridad social, si bien no incluye todo el ámbito de decisiones automatizadas en este terreno conforme al art. 130 LGSS pues, al centrarse en el acceso o retirada de servicios o prestaciones, deja al margen "los procedimientos de afiliación, cotización y recaudación". Hay que entender, por otro lado, que extiende su ámbito de aplicación más allá de lo que venimos considerando gestión de la seguridad social, toda vez que concesión, reducción o retirada de dichos servicios y prestaciones o la reclamación de su devolución son también objeto de actuaciones sancionadoras en los términos de los arts. 17, 24 ss. y 47 LISOS. En fin, la noción

de IA que maneja el Reglamento impone, como se desprende de las consideraciones desarrolladas más arriba, que las herramientas de IA predictiva que se empleen en este terreno, pese a no dar origen a decisiones automatizadas –o a "resoluciones", conforme a la nomenclatura empleada por el art. 130.1 LGSS– se sujeten a sus prescripciones.

En cuanto a las restantes cuestiones vinculadas al poder administrativo sancionador en materia social no quedan afectadas por el Anexo III. Es verdad que su apartado 6 se dedica a los sistemas de IA relacionados con la "garantía del cumplimiento del Derecho, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable". Sin embargo, el Reglamento de Inteligencia Artificial está pensando exclusivamente en las cuestiones penales. Se parte de que "las actuaciones de las autoridades garantes del cumplimiento del Derecho que implican determinados usos de los sistemas de IA se caracterizan por un importante desequilibrio de poder y pueden dar lugar a la vigilancia, la detención o la privación de libertad de una persona física, así como tener otros efectos negativos sobre los derechos fundamentales consagrados en la Carta" (§ 59). Y, sobre esta base, las definiciones normativas se centran en aquellas. En esta línea, el art. 3.45) RIA define la «autoridad garante del cumplimiento del Derecho» como "a) toda autoridad pública competente para la prevención, la investigación, la detección o el enjuiciamiento de delitos o la ejecución de sanciones penales, incluidas la protección frente a amenazas para la seguridad pública y la prevención de dichas amenazas, o b) cualquier otro organismo o entidad a quien el Derecho del Estado miembro haya confiado el ejercicio de la autoridad pública y las competencias públicas a efectos de prevención, investigación, detección o enjuiciamiento de delitos o ejecución de sanciones penales, incluidas la protección frente a amenazas para la seguridad pública y la prevención de dichas amenazas". En esta misma línea la «garantía del cumplimiento del Derecho» es definida en el apartado 46) como "las actividades realizadas por las autoridades garantes del cumplimiento del Derecho, o en su nombre, para la prevención, la investigación, la detección o el enjuiciamiento de delitos o la ejecución de sanciones penales, incluidas la protección frente a amenazas para la seguridad pública y la prevención de dichas amenazas". Por lo demás, el apartado 6 del Anexo III ni siquiera contempla todos los aspectos penales: solo ciertos aspectos que despliegan efectos particularmente intensos sobre las personas. A la postre, el uso administrativo de la inteligencia artificial no parece quedar afectado por el nuevo Reglamento.

Diferente es, en fin, el caso del uso judicial de la inteligencia artificial, al que se refiere el apartado 8 del anexo III cuya rúbrica es "Administración de justicia y procesos democráticos". Se incluyen en su apartado a) los "sistemas de IA destinados a ser utilizados por una autoridad judicial, o en su nombre, para ayudar a una autoridad judicial en la investigación e interpretación de hechos y de la ley, así como en la garantía del cumplimiento del Derecho a un conjunto concreto de hechos, o a ser utilizados de forma similar en una resolución alternativa de litigios". El uso del verbo "ayudar" implica que el RIA no está pensando en sistemas que generen decisiones judiciales automáticas pues presupone que son adoptadas por la "autoridad judicial". Estamos pues pensando en IA generativa, relacionada con la interpretación del derecho, lo que nos sitúa en el terreno

de lo que legalmente se denomina "actuaciones asistidas" (art. 57 RDL 6/2023). Desde este modo, los automatismos procedimentales o las "actuaciones proactivas" (art. 56 RDL 6/2023) no serían considerados sistemas de alto riesgo

Conviene reparar, por último, aunque no sea por ello menos importante, en que la consideración de un sistema de IA como de alto riesgo no impide su utilización; únicamente sujeta su distribución y utilización a determinadas garantías establecidas en los arts. 8 ss. Estas afectan, de un lado, a su configuración e incluyen, aparte los requisitos de "precisión, solidez y ciberseguridad" (art. 15), el establecimiento y mantenimiento de un sistema de gestión de riesgos (art. 9) –del que resulta complementaria la "conservación de registros" (art. 12)–, el control de los datos utilizados en el entrenamiento (art. 10), las obligaciones de documentación técnica (art. 11) y transparencia (art. 13) y, en fin, la "supervisión humana" (art. 14). De otra parte, comprenden específicas obligaciones para distribuidores y responsables del despliegue (arts. 16 ss.). No resulta posible, ni seguramente necesario, entrar al detalle sobre todo ello, sin perjuicio de que pueda ser traído a colación cuando sea necesario.

## 3. Una valoración de conjunto

Hecha esta reflexión sobre la regulación del RIA y su impacto en el tema que nos ocupa, es posible entrar en el último aspecto: la valoración de si nuestra regulación es adecuada o no lo es. La idea es no centrarse solo, ni tanto, en el posible impacto en ella del Reglamento de Inteligencia Artificial –que ha sido sucintamente descrita en el apartado anterior–. Pretendo más bien verificar si, del conjunto de normas examinadas, se desprende una razonable salvaguarda de los intereses de los posibles sujetos afectados. Parto, al respecto, de que la inteligencia artificial es una nueva realidad a cuya utilización debemos prestar atención, incluso cuando se utiliza por el bien común. Mi impresión general es, sin embargo, que los laboralistas hemos volcado nuestro interés fundamentalmente en el uso por las empresas de la inteligencia artificial. Aunque seguramente ello es razonable habida cuenta el objeto y los principios esenciales de nuestra disciplina, no podemos quedarnos solo en eso: hemos de fijarnos también en el uso de la IA por los organismos públicos. Si se acepta este punto de partida, se hace necesaria una revisión crítica de la situación a la que conduce el entramado de las normas europeas e internas en relación con la utilización por las entidades públicas de la inteligencia artificial que, a mi juicio, pasa por tres órdenes de consideraciones.

La primera es una crítica a la fragmentación que se desprende de la normativa aplicable. Acabamos de ver como el Reglamento de Inteligencia Artificial obliga a parcelar el tratamiento de la incorporación de los sistemas de IA a las decisiones públicas. Pero esto no parece razonable no solo por lo que está en juego desde la perspectiva ciudadana sino también por la dificultad de introducir particiones que, en rigor, distan de tener un sentido claro cuando estamos considerando el funcionamiento de instituciones públicas. Las normas del RIA, como anteriormente las del RGPD, piensan fundamentalmente en

la protección de las personas físicas; y por ello, cuando se proyectan sobre el problema que nos ocupa, dejan un amplio margen de actuaciones en el que no son aplicables. No es fácil, sin embargo, aceptar que la misma entidad quede sujeta a determinadas obligaciones solo cuando están aquellas implicadas y no lo esté en caso de que la afectación se refiera a personas jurídicas. Las entidades públicas son únicas y parece razonable que el régimen de sus actuaciones debe ser uniforme, con independencia de la naturaleza de los sujetos que queden afectados por ellas. Esto se advierte de forma bastante clara en el terreno de la utilización de la IA en la justicia, que siempre es calificada como de alto riesgo por el Reglamento, sin fijar la atención en los afectados por la actuación judicial como hemos visto en el apartado anterior. Si esta diversidad es criticable, se impone intentar una reconstrucción del tratamiento del uso de la IA que supere la diversificación normativa y que lo sujete a un marco de garantías unitario.

Esto es relativamente sencillo, en segundo lugar, en relación con algunos aspectos de las garantías a las que hemos hecho referencia. Es el caso del principio de la supervisión humana de la IA y sus derivadas, como el "derecho a de explicación de decisiones tomadas individualmente". Aunque el RIA los establezca solo para los sistemas de alto riesgo (cfr. arts. 14 y 86, respectivamente), lo cierto es que pueden deducirse derechos análogos de otras fuentes, con independencia de que nos encontremos o no en ante el funcionamiento de uno de ellos. Es verdad que las reglas sobre decisiones automatizadas del art. 22 RGPD, aunque son aplicables con independencia de que la IA que las genere sea o no de alto riesgo, tienen un ámbito de aplicación limitada en la medida en que solo las personas físicas son titulares del derecho de protección de datos (cfr. art. 1.1). Sin embargo, a través de reglas internas del máximo nivel, habrá que entender que se alcanzan resultados similares a los que derivan de una normativa europea que, en principio, no resultaría aplicable. Sin entrar en el debate general sobre si las personas jurídicas son o no titulares de los derechos fundamentales reconocidos por nuestra Constitución, parece que esta titularidad no puede discutirse de aquellos de ellos de carácter más formal o relacional, entre los que sin duda se incluyen los contenidos en el art. 24 CE. Asimismo, las exigencias constitucionales respecto al funcionamiento de los órganos administrativos (cfr. art. 105) y jurisdiccionales (art. 120) imponen específicas características a los procedimientos que conducen a sus decisiones que no pueden sino proyectarse en la configuración infraconstitucional de aquellos que se apoyen en el uso de inteligencia artificial imponiendo la supervisión humana, antes y/o después de la adopción de la decisión. De este modo, la regulación legal vigente conduce a la aplicación, por vía de las reglas sobre audiencia, motivación o recurso, de criterios de análoga significación a los derivados del principio de supervisión humana del RIA o a los establecidos al respecto en el art. 22.3 RGPD.

En esta línea, cabe observar que la incorporación de la inteligencia artificial a decisiones judiciales o administrativas tiende a no sustituir la intervención humana. En la regulación actual, esto es muy claro para las "actuaciones asistidas" en el terreno jurisdiccional que, como hemos visto, siempre quedan sujetas a la "validación" del órgano judicial (art. 57 RDL 6/2023). Pero también lo es para las actuaciones administrativas que aparecen como más polémicas desde el punto de vista de los derechos fundamentales: las

actas automatizadas de la Inspección. Al margen de que, conforme a las reglas generales, las actas implican únicamente la iniciación del procedimiento sancionador, las reglas reglamentarias aplicables imponen la intervención humana del personal de la inspección en caso de que el interesado alegue "hechos o circunstancias distintos a los consignados en el acta, insuficiencia del relato fáctico de dicho acta, o indefensión por cualquier causa" (cfr. art. 47.3 Reglamento de procedimiento sancionador en el orden social, en su redacción por RD 688/2021). Los usos predictivos de la IA, que coadyuvan a la planificación de las actuaciones inspectoras, no plantean, por lo demás, problemas relevantes desde la perspectiva del principio de supervisión humana pues se limitan a fijar el blanco para unas posteriores actuaciones inspectoras se desarrollarán por personas físicas (Todolí Signes, 2020).

Es posible, sin embargo, que ello no sea suficiente. Con carácter general se ha señalado, a propósito de los sesgos discriminatorios, la aparición de discriminaciones de nuevo cuño que, además, son mucho más sutiles que las tradicionales, por las dificultades para detectarlas y por las capacidades de justificación al alcance de la inteligencia artificial. No hay que descartar, por otro lado, el impacto del temor reverencial o la admiración que puede suscitar del funcionamiento de la IA. Es el llamado sesgo de automatización "por el que, en general, los humanos tendemos a confiar en lo que nos dicen los sistemas (de IA) empleados y rara vez lo cuestionamos" (Ponce Solé, 2024, p. 177). En este contexto, acaso la intervención humana resulte por si sola insuficiente si no va acompañada de específicas acciones de entrenamiento para mejorar; y, sobre todo, resulta necesario completar su acción mediante el establecimiento de nuevos mecanismos de control de las decisiones públicas, quizá estableciendo por ejemplo una presunción de falibilidad (Tahirí Moreno, 2023), o, en todo caso, mejorando los de carácter previo a la incorporación de la tecnología a la actividad administrativa y jurisdiccional garantizando la fiabilidad de los datos que se utilizan y haciendo transparentes las características de los procesos que desarrollan.

Diría que es este, en tercer lugar, el terreno en el que se detectan los mayores déficits, como se ha puesto de manifiesto de forma insistente (Goerlich Peset, 2021, p. 37 ss., con referencias adicionales; con posterioridad, por ejemplo, Jiménez-Castellanos Ballesteros, 2023). Pero no quedan solventados por la entrada en vigor del Reglamento. Es verdad que su art. 26.8 RIA impone también obligaciones de registro a "autoridades públicas o instituciones, órganos y organismos de la Unión" que sean "responsables del despliegue de sistemas de IA de alto riesgo" y les impide desplegar aquellos que no hayan sido registrados previamente por su proveedor o distribuidor. El alcance de la obligación de registro se concreta en los arts. 49 y 71 y supone que, con algunas salvedades que resultan de poco interés en nuestra materia, una información relativamente amplia del sistema (anexo VIII RIA) queda a disposición del público "de manera sencilla" (art. 71.4).

Estas reglas completan las establecidas por el ordenamiento interno, cuando el sistema en cuestión tiene la condición de alto riesgo. Así ocurre en el marco de las actuaciones judiciales, en las que el marco formal establecido por el art. 58 RDL 6/2023 se completa con las previsiones del RIA. De hecho, este precepto anticipa las previsiones europeas, puesto que los sistemas afectados no son solo los que presuponen decisiones automatizadas: las actuaciones proactivas en las que la decisión final es humana están tam-

bién incluidas en "la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, la auditoría del sistema de información y de su código fuente" (apartado 1) y se prevé la publicidad y objetividad de "los criterios de decisión, dejando constancia de las decisiones tomadas en cada momento" (apartado 2).

En el ámbito administrativo, sin embargo, las cosas son más complicadas, habida cuenta que, en su mayor parte, las actuaciones algorítmicas quedan, como sabemos, extramuros de la consideración de sistemas de IA de alto riesgo, con la única salvedad de su uso en la dispensación de servicios o prestaciones vinculados a la protección social. En este contexto, disponemos únicamente de las reglas generales (art. 41 Ley 40/2015) o específicas (art. 130.I LGSS), cuyas insuficiencias son notables por dos razones. La primera se relaciona con la estricta noción de decisión automatizada, que incluye únicamente la "adopción de resoluciones" en el caso de la gestión de la Seguridad Social (art. 130.I LGSS) o, si se amplía la óptica hasta incluir "cualquier acto o actuación realizada íntegramente a través de medios electrónicos" como hace el art. 41.1 40/2015, se requiere "que no haya intervenido de forma directa un empleado público". De este modo, las actuaciones prospectivas de las entidades administrativas quedan fuera de las obligaciones de transparencia legalmente establecidas. Estas, en segundo lugar, son muy limitadas. Literalmente se concretan en la determinación previa del "órgano u órganos competentes, según los casos, para la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, auditoría del sistema de información y de su código fuente" y en la información sobre "el órgano que debe ser considerado responsable a efectos de impugnación" (arts. 41.2 Ley 40/2015 y 130.II LGSS). La transparencia es, pues, estrictamente «orgánica» –quién determina las características y ante quién se pueden recurrir las actuaciones automatizadas– y en modo alguno tiene componentes «funcionales» –relacionadas con la dinámica del sistema algorítmico–.

Seguramente este estado de cosas debe ser revisado. Textos posteriores adoptan, en este sentido, una perspectiva más amplia. El art. 23 Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, se sitúa en ella al referirse de forma genérica a "los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas", expresión notablemente más amplia que las empleadas en las normas anteriores. Asimismo, hace específicas alusiones a la transparencia. Más allá de su consideración en el art. 23.1, se prevé que las administraciones públicas "priorizarán la transparencia en el diseño y la implementación y la capacidad de interpretación de las decisiones adoptadas por los mismos". Poco antes, la Carta de Derechos Digitales, aprobada por el Gobierno en julio de 2021, había insistido en esta línea. En su apartado XVII.6, en el marco de la promoción de "los derechos de la ciudadanía en relación con la inteligencia artificial… en el marco de la actuación administrativa" se reconocieron, entre otros, los derechos a "que las decisiones y actividades en el entorno digital respeten los principios de buen gobierno y el derecho a una buena Administración digital, así como los principios éticos que guían el diseño y los usos de la inteligencia artificial" (letra a]) y, sobre todo, a "la transparencia sobre el uso de instrumentos de inteligencia artificial y sobre su funcionamiento y alcance en cada procedimiento concreto y, en particular,

acerca de los datos utilizados, su margen de error, su ámbito de aplicación y su carácter decisorio o no decisorio" (letra b]).

Sin embargo, desde una perspectiva práctica estos textos tienen una virtualidad limitada. La Carta de Derechos Digitales "no tiene carácter normativo". Así se reconoce de forma taxativa en sus "consideraciones previas". Por ello, y en relación con los derechos de transparencia aludidos se prevé que "la ley podrá regular las condiciones de transparencia y el acceso al código fuente, especialmente con objeto de verificar que no produce resultados discriminatorios"; y en este nivel jerárquico estamos donde estábamos. Por su parte, el citado art. 23 Ley 15/2022, como muchos de los preceptos de esta norma, tienen un claro componente programático que dificulta extraer concretas prescripciones de él. Por lo demás, una y otro se centran en el terreno de los derechos fundamentales lo que incide en la discutible fragmentación del tratamiento del uso de la IA por las Administraciones, en función de que sean afectadas personas físicas o jurídicas.

A la espera de avances normativos específicos y más concretos, acaso sea posible superar este estado de cosas mediante el recurso a la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno. El argumento que suministra esta norma en relación con el problema que nos ocupa se encuentra en su art. 7, que en el marco de la "transparencia de la actividad pública" regula la "información de relevancia jurídica". En concreto, obliga a las Administraciones Públicas, en el ámbito de sus competencias, a publicar "las directrices, instrucciones, acuerdos, circulares o respuestas a consultas planteadas por los particulares u otros órganos en la medida en que supongan una interpretación del Derecho o tengan efectos jurídicos".

Existen ya supuestos de aplicación por el Consejo de Transparencia y, en último término, por los órganos judiciales, de esta regla al ámbito que nos interesa. Por lo que se refiere a los primeros, los ha estudiado recientemente Amparo García Rubio (2024, Parte 3.2). A su juicio, las resoluciones del indicado Consejo, que han estimado las reclamaciones frente a las solicitudes de información frente a las entidades gestoras o la ITSS, han de llevar a la obvia conclusión de que "entre la nada y el todo existen niveles intermedios de información que, teniendo en cuenta las particularidades y fines de la actuación inspectora, seguramente proporcionarán ese punto de equilibrio a lograr entre la eficacia administrativa y la transparencia algorítmica". Cabe indicar que, en otros ámbitos, estas cuestiones han desembocado ya en procesos judiciales en los que se ha buscado ir más allá de la posición del Consejo. Aparte la STSJ Galicia cont. 89/2024, 22 marzo, en relación con el algoritmo de un programa de citación de usuarios en el sistema público sanitario, es necesario traer a colación las resoluciones recaídas en el llamado caso Bosco, en el que, a instancias de una fundación privada, se analiza la información que debe ser suministrada en relación con el algoritmo que controla la asignación del bono social. Sin entrar en detalles sobre las distintas resoluciones recaídas (Res. Consejo Transparencia R/0701/2018 y sentencias JC-8 cont. 143/2021, 30 diciembre y AN cont. 30 abril 2024, rec. 51/2022), este episodio abrirá probablemente la posibilidad de que el Tribunal Supremo se pronuncia sobre su alcance y, en concreto, sobre las posibilidades de acceder al conocimiento del código-fuente. Con toda probabilidad, y con independencia

de la solución que se dé a esta cuestión, parece necesaria una intervención normativa ulterior, que implique la adecuación por parte de las Administraciones a la doctrina que va siendo emanada del Consejo de Transparencia y haga efectivos los principios que se desprenden de las citadas fuentes de *soft law*.

## Bibliografía citada


Aibar Bernard, J. (2020). El Big Data y el análisis de datos aplicados por la Tesorería General de la Seguridad Social como medio de lucha contra el fraude en la Seguridad Social. *Trabajo y derecho: nueva revista de actualidad y relaciones laborales*, *11*.

AISS. (2019). *10 desafíos mundiales para la Seguridad Social. Evolución e innovación*. https://www.issa.int/sites/default/files/documents/publications/3-10-challenges-Global-2019-WEB-263632.pdf.

García Rubio, M. A. (2024). Retos actuales de la Inspección de Trabajo y Seguridad Social desde la perspectiva de los derechos de los administrados. *Documentación-Laboral*, *II*(132).

Goerlich Peset, J. M. (2021). Decisiones administrativas automatizadas en materia social: algoritmos en la gestión de la Seguridad Social y en el procedimiento sancionador. *Labos: Revista de Derecho del Trabajo y Protección Social*, *2*(2), 22-42.

Jiménez-Castellanos Ballesteros, I. (2023). Decisiones automatizadas y transparencia administrativa: Nuevos retos para los derechos fundamentales. *Revista española de la transparencia*, *16*, 191-215.

Nores Torres, L. E. (2023). El proceso de digitalización en la jurisdicción social: algunos avances y perspectivas. *Lex social: revista de los derechos sociales*, *13*(2).

Nores Torres, L. E. (2024a). Digitalización, inteligencia artificial y proceso laboral. *Labos: Revista de Derecho del Trabajo y Protección Social*, *5*(1), 222-246.

Nores Torres, L. E. (2024b). *Proceso laboral y digitalización tras el RDL 6/2023, de 19 de diciembre*. Tirant lo Blanch.

Ponce Solé, J. (2024). Inteligencia Artificial, decisiones administrativas discrecionales totalmente automatizadas y alcance del control judicial: ¿indiferencia, insuficiencia o deferencia? *Revista de Derecho Público: teoría y método*, *9*, 171-220.

Redondo Rincón, G. (2020). Las nuevas tecnologías en el control de la Incapacidad Temporal: la aplicación de la analítica predictiva. *Trabajo y derecho: nueva revista de actualidad y relaciones laborales*, *11*.

Tahirí Moreno, J. (2023). El principio de presunción de falibilidad de las decisiones algorítmicas desfavorables: Una nueva garantía jurídica frente a las decisiones automatizadas y el uso de sistemas de inteligencia artificial en la administración pública. *Revista Aragonesa de Administración Pública*, *60*, 188-214.

Todolí Signes, A. (2020). Retos legais do uso do big data na selección de suxeitos a investigar pola Inspección de Traballo e da Seguridade Social. *REGAP: Revista galega de administración pública*, *1*(59), 79-102.