

INTELIGENCIA ARTIFICIAL, GÉNERO Y TRABAJO: PROPUESTAS JURÍDICAS PARA LA PROTECCIÓN DE LA IGUALDAD Y LA NO DISCRIMINACIÓN

MARÍA JOSÉ ROMERO RODENAS

Con la participación de:

DELKA EDUARDA ÁNGELES RODRÍGUEZ

**INTELIGENCIA ARTIFICIAL, GÉNERO Y TRABAJO:
PROPUESTAS JURÍDICAS PARA LA PROTECCIÓN DE LA
IGUALDAD Y LA NO DISCRIMINACIÓN**

María José Romero Rodenas

*Catedrática de Derecho del Trabajo y Seguridad Social.
Universidad de Castilla-La Mancha.*

Con la participación de:

Delka Eduarda Ángeles Rodríguez

*Contratada Predoctoral para la Formación de Personal
Investigador en Derecho del Trabajo y Seguridad Social.
Universidad de Castilla-La Mancha.*

Esta obra se enmarca en el Proyecto de Investigación concedido mediante Resolución de 24/09/2025, del Instituto de la Mujer de Castilla-La Mancha, por la que se conceden ayudas a la investigación en el año 2025 (DOCM 8 de octubre de 2025).

El contenido del mismo es responsabilidad exclusiva de las autoras y no refleja necesariamente la posición oficial del Instituto de la Mujer de Castilla-La Mancha.

Primera edición: 2026

© Editorial Bomarzo S.L.
c/ Dionisio Guardiola 1, 4º
02002 Albacete (España)
editorialbomarzo@editorialbomarzo.es
www.editorialbomarzo.es

Quedan rigurosamente prohibidas, sin la autorización escrita del titular del "Copyright", bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler.

Impreso en España.

ISBN: 979-13-88063-08-4

DL: AB 172-2026

Diseño de la colección: Javier Hidalgo Romero.

Maquetación: Juan Gómez Rubio.

Imprime: Veprix S.L.

ÍNDICE

PRESENTACIÓN	9
1. INTRODUCCIÓN	15
2. MARCO JURÍDICO DE LA IA LABORAL: DATOS SENSIBLES, SUPERVISIÓN Y LÍMITES DEL SECRETO EMPRESARIAL.....	23
3. DISCRIMINACIÓN ALGORÍTMICA Y “PROBLEMA DEL PROXY” EN EL REGLAMENTO EUROPEO DE INTELIGENCIA ARTIFICIAL	35
3.1. La documentación técnica, la trazabilidad y la gobernanza de los datos en el Reglamento de Inteligencia Artificial: alcance y límite para la prevención de los sesgos de género	38
3.2. Calidad de los datos, tratamiento de categorías sensibles y sesgos de género en el art. 10 del Reglamento de Inteligencia Artificial.....	45
4. DISCRIMINACIÓN ALGORÍTMICA	51
4.1. Datos que reflejan discriminación algorítmica.....	51
4.2. Vulneración de derechos fundamentales: derecho a la no discriminación	60
4.3 Opacidad algorítmica, trazabilidad del “pipeline” y riesgos de discriminación	64
4.4. Gobernanza de datos en la IA del alto riesgo: de la “IA centrada en el modelo” a la “IA centrada en los datos” para prevenir sesgos y discriminación	70
4.5 Entre “fairness” y no discriminación: límites del Reglamento de IA para proteger el derecho a la igualdad	73
4.6 Derecho a la explicación como llave para denunciar la discriminación algorítmica	77

4.7 Sanción y reparación frente a la discriminación algorítmica: entre la tutela y las nuevas facultades del Reglamento de IA.....	79
---	----

5. EXPERIENCIAS COMPARADAS EN LA REGULACIÓN DE LA IA LABORAL CON PERSPECTIVA DE IGUALDAD.....	85
--	----

5.1. Regulaciones subnacionales en Estados Unidos sobre algoritmos de selección y auditorías de sesgo	85
--	----

5.2. Intervención de autoridades laborales y antidiscriminatorias en el uso de IA con enfoque de igualdad.....	89
---	----

6. GOBERNANZA FEMINISTA DE LA IA EN LA EMPRESA: MÁS ALLÁ DEL CUMPLIMIENTO TÉCNICO. PROPUESTAS	93
--	----

6.1. Incorporar la igualdad de género en la política algorítmica corporativa: propuestas	94
---	----

6.2. Reforzar la composición y funciones de los órganos internos de decisión: propuestas	94
---	----

6.3. Transparencia interna y rendición de cuentas con enfoque de género: propuestas.....	95
---	----

7. IGUALDAD Y NEGOCIACIÓN COLECTIVA ANTE LA IA: CLÁUSULAS DE NO DISCRIMINACIÓN ALGORÍTMICA. PROPUESTAS	97
--	----

7.1. Cláusulas tipo sobre información y consulta: propuestas	98
--	----

7.2. Comisiones paritarias de seguimiento algorítmico: propuestas	99
---	----

7.3. Derechos individuales de revisión y explicabilidad laboral: propuestas.....	99
---	----

8. IA, CUIDADOS Y ORGANIZACIÓN DEL TIEMPO DE TRABAJO: RIESGOS DE DISCRIMINACIÓN INDIRECTA. PROPUESTAS	101
---	-----

8.1. Evaluaciones de impacto de género antes de automatizar la organización del tiempo de trabajo: propuestas	102
--	-----

8.2. Revisión de criterios de “idoneidad” y “rendimiento”: propuestas.....	103
8.3. Integrar la IA en los planes de igualdad y en los diagnósticos de brecha: propuestas.....	103
 9. AUTORIDADES DE CONTROL Y PERSPECTIVA DE GÉNERO: HACIA UNA SUPERVISIÓN PROACTIVA DE LA IA LABORAL.....	105
9.1. Mandatos explícitos de igualdad para autoridades de IA, de datos y de igualdad: propuestas	106
9.2. Protocolos de supervisión con indicadores de género: propuestas.....	107
9.3. Uso estratégico del grupo de expertos científicos: propuestas.....	107
 10. CONCLUSIONES	109
10.1. La IA en el trabajo no es neutra: reproduce y puede amplificar las desigualdades de género.....	109
10.2. El marco normativo vigente es necesario pero insuficiente frente a la discriminación algorítmica de género	110
10.3. La gobernanza de la IA debe feminizarse: igualdad sustantiva como criterio rector de diseño y uso.....	111
10.4. La negociación colectiva es un instrumento clave para construir garantías específicas frente a la discriminación algorítmica.....	111
10.5. IA, cuidados y organización del tiempo de trabajo: el núcleo de la discriminación indirecta.....	112
10.6. Las autoridades de control deben asumir un mandato explícito de igualdad en la supervisión de la IA.....	113
10.7. Hacia un marco integrado de “derecho del trabajo algorítmico” con perspectiva de género	114
 11. ANEXO BIBLIOGRAFÍA	117

PRESENTACIÓN

La irrupción de la inteligencia artificial en las relaciones laborales constituye uno de los procesos de transformación más profundos del Derecho del Trabajo contemporáneo, pero también uno de los más ambivalentes desde la óptica de la igualdad entre mujeres y hombres. Los sistemas algorítmicos que hoy intervienen en la selección, contratación, asignación de tareas, evaluación del desempeño o incluso la extinción de los contratos se presentan retóricamente como herramientas “objetivas” y “neutrales”, capaces de reducir la discrecionalidad humana y aumentar la eficiencia en la toma de decisiones. Sin embargo, cuando se analizan con detenimiento sus condiciones de diseño, los datos que los alimentan y los contextos en los que operan, emerge un panorama mucho más complejo: lejos de ser inmunes a los sesgos, estos sistemas pueden reproducir, amplificar y sofisticar las desigualdades de género ya arraigadas en el mercado de trabajo.

Este trabajo parte precisamente de esa constatación: la inteligencia artificial no es una tecnología neutra, ni en términos sociales ni, mucho menos, en términos de género. Los modelos de aprendizaje automático aprenden de datos históricos que reflejan prácticas discriminatorias, segregación ocupacional, techos de cristal y penalizaciones ligadas a la maternidad y a los cuidados, todo ello en un entorno en el que las mujeres siguen estando infrarrepresentadas en los sectores tecnológicos y en los espacios donde se toman las decisiones sobre diseño y gobernanza de la IA. Cuando estos datos sesgados se convierten en materia prima para entrenar algoritmos destinados a filtrar candidaturas, recomen-

dar ascensos o valorar el rendimiento, la desigualdad deja de ser solo una herencia del pasado para convertirse en una característica estructural del futuro del trabajo si no se establecen contrapesos normativos robustos.

Desde una perspectiva jurídica y con enfoque de género, el presente estudio se propone desentrañar cómo interactúan los sistemas de inteligencia artificial con el derecho fundamental a la igualdad y a la no discriminación de las personas trabajadoras, en particular de las mujeres. Ello exige abandonar una mirada puramente tecnocéntrica para situar la IA en el entramado de relaciones de poder que atraviesan la empresa y el mercado de trabajo, donde las decisiones aparentemente técnicas tienen consecuencias muy concretas sobre quién accede al empleo, qué trayectorias profesionales se reconocen como “meritorias” y cómo se distribuyen las oportunidades de promoción, estabilidad y protección social. Esta investigación asume, por tanto, que los algoritmos son nuevos dispositivos de poder empresarial y que su análisis ha de hacerse desde el prisma del Derecho del Trabajo y de la Seguridad Social, con especial atención a los impactos diferenciales que producen sobre las mujeres y otros colectivos protegidos.

Para ello, el estudio se articula en torno a un eje claramente definido: la construcción de un marco jurídico de la inteligencia artificial que incorpore de manera explícita y sistemática la perspectiva de género. En primer lugar, se examina el contexto normativo en el que se inserta la IA aplicada al trabajo, comenzando por la definición y alcance de la inteligencia artificial y los algoritmos, y siguiendo por el análisis del Reglamento Europeo de Inteligencia Artificial y su interacción con el Reglamento General de Protección de Datos, las directivas europeas antidiscriminatorias y la normativa española en materia de igualdad y no discriminación. Este recorrido permite mostrar que el derecho vigente, aunque ha avanzado en la identificación de riesgos y en la imposición de ciertas garantías, todavía presenta lagunas significativas cuando se enfrenta a la discriminación algorítmica de género.

Un núcleo central de la investigación lo constituye el análisis sistemático de las obligaciones que el Reglamento Europeo de Inteligencia Artificial impone a los sistemas de alto riesgo –entre ellos, los destinados a la contratación, gestión y supervisión de personas trabajadoras– y su potencial para afrontar los sesgos de género. Se estudian en profundidad el sistema de gestión de riesgos, las exigencias de gobernanza de datos, la documentación técnica, la transparencia, la supervisión humana y los requisitos de calidad de los datos, con especial atención al artículo 10 y a la posibilidad de tratar categorías especiales de datos personales para detectar y corregir sesgos. Desde una lectura con perspectiva de género, se pone de relieve que estas obligaciones constituyen un avance relevante, pero no suficiente, en la medida en que se centran en la corrección técnica del sesgo sin garantizar plenamente la efectividad del derecho a la igualdad sustantiva tal como lo ha configurado el derecho antidiscriminatorio europeo y español.

La investigación dedica, además, un capítulo específico a la discriminación algorítmica y a sus manifestaciones en el ámbito laboral, distinguiendo cuidadosamente entre sesgo y discriminación, y subrayando que el cumplimiento formal de las obligaciones técnicas del Reglamento de IA no equivale automáticamente al respeto del derecho fundamental a la no discriminación. Se muestran distintos escenarios en los que el uso de proxies, la falta de representatividad de los datos o decisiones de diseño aparentemente neutras terminan generando efectos adversos desproporcionados sobre las mujeres, especialmente en procesos de selección, clasificación de perfiles, evaluación del desempeño u organización del tiempo de trabajo. Asimismo, se analizan las limitaciones de las métricas de equidad actualmente disponibles, muchas de ellas orientadas a una igualdad meramente formal, y su insuficiente alineación con el ideal de igualdad sustantiva que inspira la jurisprudencia del Tribunal de Justicia de la Unión Europea y la normativa específica de igualdad de trato.

Otro de los aportes relevantes de la investigación reside en el examen de los mecanismos de transparencia y aplicabilidad de los sistemas de inteligencia artificial, así como de su relevancia para la protección procesal del derecho a la igualdad de las personas trabajadoras. En este sentido, el estudio analiza el derecho previsto en el Reglamento de IA, poniendo en diálogo sus exigencias con las obligaciones de información y las garantías ya presentes en el RGPD y en la normativa laboral española, en particular en lo relativo al uso de decisiones automatizadas y elaboración de perfiles. Desde una óptica de género, se enfatiza que el derecho a comprender el funcionamiento de los sistemas que afectan a los derechos laborales no es un lujo técnico, sino un requisito indispensable para que las mujeres puedan identificar y denunciar prácticas discriminatorias, superar la opacidad algorítmica y acceder a una tutela judicial efectiva.

La dimensión reparadora y sancionadora de la igualdad ocupa también un espacio central en este trabajo. Se estudian los mecanismos existentes en la jurisdicción social para la reparación del daño derivado de la discriminación, tanto en la normativa general (Ley 36/2011, de 10 de octubre, Reguladora de la Jurisdicción Social y Real Decreto Legislativo 5/2000, de 4 de agosto, por el que se aprueba el Texto Refundido de la Ley sobre Infracciones y Sanciones en el Orden Social) como en la legislación específica en materia de igualdad y no discriminación. El análisis se detiene en cuestiones clave como la inversión de la carga de la prueba, la presunción de daño moral en casos de discriminación acreditada, la cuantificación de las indemnizaciones y las exigencias de efectividad, proporcionalidad y carácter disuasorio de las sanciones, examinando cómo estos elementos se ven tensionados cuando la discriminación tiene su origen en sistemas de IA cuya lógica interna resulta difícil de desentrañar.

La investigación no se limita al plano institucional o judicial, sino que incorpora la dimensión colectiva y participativa de la regulación de la inteligencia artificial en la empresa, con numerosas pro-

puestas prácticas a implantar en la empresa. Desde una perspectiva de género, se subraya la importancia de reforzar el papel de la representación legal de las personas trabajadoras en el acceso a la información sobre sistemas algorítmicos, en la negociación de garantías específicas frente a la discriminación y en la construcción de modelos de auditoría que integren indicadores de impacto de género. Se propone entender la negociación colectiva como un espacio privilegiado para pactar mecanismos de transparencia reforzada, protocolos de evaluación de sesgos, cláusulas de no discriminación algorítmica y medidas de corrección que vayan más allá del cumplimiento mínimo normativo, contribuyendo a transformar las estructuras organizativas que sostienen la desigualdad.

El análisis también mira hacia la dimensión europea e institucional, analizando el papel del grupo de expertos científicos independientes en inteligencia artificial, la Oficina Europea de IA y las autoridades nacionales de vigilancia del mercado como actores clave en la identificación de riesgos para la igualdad y en la supervisión de sistemas de alto riesgo con potencial impacto discriminatorio. Se examinan los instrumentos de los que disponen estas instancias para solicitar información, realizar evaluaciones técnicas, orientar las prácticas de auditoría y articular respuestas coordinadas frente a vulneraciones del derecho de igualdad de trato, insistiendo en la necesidad de que la perspectiva de género forme parte explícita de sus mandatos, metodologías y prioridades de actuación.

En suma, la investigación propone un marco interpretativo y propositivo en el que la inteligencia artificial en el trabajo se analiza no solo como un desafío técnico-regulatorio, sino como un campo de disputa en torno a la igualdad de género y a la garantía de los derechos fundamentales de las personas trabajadoras. Al combinar un análisis jurídico exhaustivo con una mirada feminista atenta a las desigualdades estructurales, el estudio aspira a ofrecer herramientas conceptuales y normativas para que legisladores, tribunales, autoridades de control, agentes sociales y co-

munidad académica puedan abordar de forma más consciente y exigente los riesgos de discriminación algorítmica. La tesis que lo recorre es clara: la transición digital del trabajo solo será legítima si se acompaña de un compromiso efectivo con la igualdad sustantiva entre mujeres y hombres; y la regulación de la inteligencia artificial, lejos de ser un ámbito neutro, constituye uno de los escenarios decisivos en los que ese compromiso ha de concretarse.

1. INTRODUCCIÓN

La inteligencia artificial no es inmune a la discriminación algorítmica y a los sesgos de género, e incluso puede ampliar la brecha de género que ya existe en el ámbito laboral cuando se incorpora sin salvaguardas específicas. Esta potencial ampliación no se limita a reproducir desigualdades conocidas, sino que puede reforzarlas bajo una apariencia de neutralidad técnica, dificultando su identificación y cuestionamiento. En la práctica, decisiones cruciales sobre acceso al empleo, estabilidad contractual, promoción profesional o evaluación del rendimiento pueden verse condicionadas por modelos que aprenden de datos históricamente sesgados¹ y que se despliegan en organizaciones donde las estructuras patriarcales siguen presentes.

Podemos resumir grandes desafíos que plantean los algoritmos de inteligencia artificial para la igualdad de género, entre los que tenemos, en primer lugar, el factor humano: las personas desarrolladoras o diseñadoras de sistemas aportan de manera consciente o inconsciente estereotipos y sesgos que se filtran en el propio diseño del modelo, en la selección de variables, en la definición de lo que se considera “mérito” o “idoneidad” y en la fijación de umbrales de decisión. Estas decisiones de diseño pueden traducir en términos matemáticos prejuicios tradicionales², sobre la

¹ FLORES ANARTE, L., “Sesgos de género en la inteligencia artificial: El Estado de derecho frente a la discriminación algorítmica por razón de sexo”, *Revista internacional de pensamiento político*, Núm. 18, 2023, pp. 95-120.

² ALONSO BETANZOS, A., “Inteligencia Artificial y sesgos de género”, *Gender on Digital*, Vol. 1 (2023), pp. 27-29. <https://doi.org/10.35869/god.v1i.5060>

menor disponibilidad de las mujeres, su supuesta menor capacidad de liderazgo o su vinculación prioritaria con los cuidados, de modo que el sistema aprende a penalizar trayectorias laborales marcadas por interrupciones, jornadas reducidas o cambios de ocupación que responden a la división sexual del trabajo. El factor humano incluye también la composición de los equipos técnicos: la infrarrepresentación de mujeres y de perfiles con formación en igualdad de género en los espacios donde se diseñan y validan los modelos reduce la probabilidad de que se detecten y corrijan estos sesgos de forma temprana.

En segundo lugar, los datos de alimentación constituyen una fuente estructural de riesgo: los sistemas aprenden a partir de información que suele contener sesgos acumulados, como evaluaciones laborales sistemáticamente más negativas para las mujeres, menores tasas de promoción, salarios inferiores o menor presencia en puestos de responsabilidad, que el modelo interpreta como correlaciones verdaderas y aplicables. Incluso cuando se eliminan explícitamente variables sensibles como el sexo, el modelo puede reconstruir esas diferencias mediante proxies (por ejemplo, patrones de dedicación temporal, interrupciones por cuidados o determinados itinerarios formativos), generando resultados que perpetúan la discriminación indirecta. A ello se añaden problemas de infrarrepresentación o sobrerrepresentación de determinados grupos: si las mujeres están infrarrepresentadas en los datos de entrenamiento, el sistema aprende peor sus patrones reales y produce predicciones menos fiables para ellas; si determinados perfiles masculinos dominan el conjunto de datos, las métricas globales de rendimiento pueden ocultar un comportamiento mucho peor respecto de las mujeres.

Por último, la transparencia y la aplicabilidad se convierten en un desafío transversal: la dificultad de entender cómo toman decisiones los modelos de aprendizaje automático complica la detección, la acreditación y la demostración procesal de la discriminación de género. La denominada “caja negra” de los sistemas complejos

—en especial redes neuronales profundas— hace que incluso quienes despliegan la IA tengan problemas para explicar qué factores han resultado determinantes en una decisión concreta, de modo que las mujeres afectadas por un trato desfavorable encuentran enormes obstáculos para reconstruir la cadena causal entre la salida del sistema y su pertenencia a un grupo protegido. Aunque el desarrollo técnico ha avanzado en métodos de aplicabilidad, no existen todavía estándares universalmente aceptados que permitan garantizar explicaciones suficientemente fieles al modelo, comprensibles para personas no expertas y útiles desde el punto de vista probatorio en procesos por discriminación.

La definición jurídica de inteligencia artificial se encuentra recogida en el Reglamento Europeo sobre Inteligencia Artificial, que opta por una concepción amplia destinada a abarcar la diversidad de técnicas y usos existentes, pero que al mismo tiempo necesita ser concretada en clave de derechos fundamentales³. Desde distintas perspectivas, la inteligencia artificial puede entenderse como modelos matemáticos diseñados para resolver problemas humanos o como sistemas capaces de interpretar información externa, aprender de ella y ejecutar tareas concretas, de forma más o menos autónoma. En su aproximación a la emulación del pensamiento humano, la IA se describe también como sistemas que detectan, actúan y aprenden; en términos prácticos, son programas que analizan grandes volúmenes de datos para automatizar procesos, sustituyendo o asistiendo decisiones que antes se tomaban de manera exclusivamente humana. Esta capacidad de procesar en segundos cantidades de información inabarcables para una persona, y de extraer patrones estadísticos ocultos, constituye su principal atractivo, pero también la raíz de nuevos riesgos de discriminación estructural si no se articula un marco de garantías adecuado.

³ MERCADER UGINA J., VELASCO PARDO B.: “El impacto en lo laboral del reglamento de Inteligencia Artificial”, *Revista Derecho Social y Empresa*, núm. 23 (2025), pp. 106-133.

En la práctica, la variante de IA que está teniendo un impacto más intenso en el empleo es la ligada al aprendizaje automático, que identifica patrones en conjuntos masivos de datos y genera predicciones de alta precisión sobre comportamientos futuros, probabilidades de éxito o riesgo asociado a determinados perfiles. Este tipo de modelos alimenta ya un amplio mercado comercial y empresarial, desde herramientas de filtrado de currículos y plataformas de gestión de recursos humanos hasta sistemas de evaluación continua del desempeño o de asignación dinámica de tareas. Su atractivo radica en ofrecer asistencia automatizada a decisiones con costes reducidos y alto rendimiento, lo que resulta especialmente tentador en escenarios de presión competitiva y búsqueda de eficiencia; sin embargo, plantea interrogantes de enorme calado en términos de transparencia, rendición de cuentas y respeto del derecho a la igualdad y a la no discriminación⁴

El derecho digital del trabajo⁵ se apoya ahora en nuevos pilares, entre los que figuran de manera destacada los algoritmos y la inteligencia artificial, que se integran en el poder de dirección empresarial y reconfiguran su ejercicio. Según la Real Academia Española, un algoritmo es “un conjunto ordenado y finito de operaciones que permiten hallar la solución a un problema”; en la práctica, esto se traduce en instrucciones matemáticas que persiguen un resultado, codificando reglas, criterios de decisión y objetivos de optimización. Más allá de la definición, lo relevante es comprender su finalidad y su lógica: los algoritmos emplean

⁴ MERCADER UGUINA J.R. “El derecho de la Unión Europea ante el desarrollo tecnológico: el reglamento de inteligencia artificial”, *Documentación Laboral* núm. 135 (2025) pp. 37-51.

⁵ Lecturas, siempre obligadas de ROJO TORRECILLAS E.: “Nuevas aportaciones sobre el impacto de la inteligencia artificial y la gestión algorítmica, en general, y en el mundo del trabajo en particular” <http://www.eduardorjotorrecilla.es/2026/02/nuevas-aportaciones-sobre-el-impacto-de.html>; Sobre tecnología, inteligencia artificial y gestión algorítmica. Impacto en el mundo del trabajo. Presentación, y ordenación y recopilación de artículos publicados en el blog (2020-2026), en http://www.eduardorjotorrecilla.es/2026/02/sobre-tecnologia-inteligencia_11.html

grandes cantidades de datos históricos para identificar relaciones y producir predicciones o aproximaciones que sustituyen juicios intuitivos por respuestas aparentemente más objetivas, facilitando decisiones automáticas o asistidas por humanos en procesos críticos como contratación, evaluación o despido. Desde la perspectiva de género, esta sustitución de juicios humanos por resultados algorítmicos no elimina la subjetividad, sino que la traslada y la encapsula en estructuras técnicas menos visibles y, por tanto, más difíciles de cuestionar.

Fuentes como la Carta Ética Europea sobre el uso de la IA en sistemas judiciales⁶ definen el algoritmo como una secuencia finita de reglas formales que transforma entradas en resultados, a menudo integradas en procesos automatizados y basadas en modelos de aprendizaje automático. El problema para el derecho⁷ es la ausencia de una definición jurídica única y operativa de “algoritmo”, lo que complica su regulación y aplicación práctica en ámbitos como el laboral, donde es preciso determinar con claridad quién responde por las decisiones automatizadas y bajo qué estándares. Esta desconexión entre herramientas de programación y categorías jurídicas se agrava cuando se aborda la igualdad y la no discriminación, pues las técnicas informáticas para corregir sesgos tienen límites evidentes y no siempre se corresponden con los conceptos jurídicos de discriminación directa, indirecta o acoso por razón de sexo. El riesgo es que los desarrollos técnicos de mitigación del sesgo se usen como argumento de cumplimiento (“el sistema ya ha sido “debiased”) sin que ello garantice, desde el punto de vista normativo, la ausencia de trato discriminatorio hacia las mujeres.

El aprendizaje automático, rama de la IA, opera mediante distintos modelos y funciones, que comparten la idea básica de apren-

⁶ Adoptada por el CEPEJ durante su 31.ª Reunión plenaria (Estrasburgo, 3-4 de diciembre de 2018).

⁷ COTINO HUESO L.: “Cómo abordar jurídicamente el impacto de la inteligencia artificial en los derechos fundamentales”, en *Derecho y Tecnologías* (Dir. Casas Baamonde y Pérez del Prado), Editorial Universitaria Ramón Areces, 2025 pp. 142 y ss.

der a partir de ejemplos previos, estableciendo relaciones entre entradas y salidas para generalizar y predecir sobre nuevos datos. Las técnicas incluyen aprendizaje supervisado, no supervisado, semisupervisado y por refuerzo, cada una con implicaciones diferentes en términos de control, interpretabilidad y riesgo de sesgo. Cuando estas técnicas se combinan en arquitecturas profundas hablamos de aprendizaje profundo: estructuras en capas que se autoajustan y se retroalimentan, integrándose en redes neuronales que optimizan su rendimiento a medida que procesan más datos. Además, el aprendizaje profundo puede complementarse con refuerzo para entrenar modelos que seleccionen opciones que maximicen un objetivo dado, como la “eficiencia” en la selección de candidaturas o la reducción de tasas de rotación, objetivos que pueden entrar en tensión con la igualdad de género si no se definen cuidadosamente.

No obstante, los modelos de aprendizaje automático presentan tres problemas principales: su diseño puede incorporar decisiones políticas y organizativas de sus creadores y, por tanto, sesgar la representación de la realidad; los datos de entrenamiento pueden ser insuficientes, sesgados o mal interpretados; y la opacidad técnica, frecuentemente por código cerrado o por complejidad estructural, dificulta entender cómo se llega a una decisión. En consecuencia, estos sistemas correlacionan datos sin establecer causalidades, lo que complica rastrear el origen de una decisión y aumenta el riesgo de predicciones erróneas o discriminatorias, especialmente cuando las correlaciones reflejan desigualdades de género históricas. El aprendizaje supervisado, por ejemplo, busca generalizar patrones a partir de datos etiquetados para predecir correctamente sobre casos nuevos, pero su fiabilidad depende de la calidad y representatividad de esos datos, de los criterios de etiquetado y de la forma en que se definen las categorías de éxito o fracaso, ámbitos donde los sesgos de género pueden introducirse fácilmente.

En suma, los algoritmos ejercen hoy un poder real sobre la sociedad y sobre las decisiones individuales, y la implantación de

la inteligencia artificial en el ámbito laboral resulta profundamente disruptiva para el ordenamiento jurídico, especialmente en lo relativo al derecho a la igualdad y a la no discriminación. La capacidad de la IA para procesar datos personales, perfilar a las personas trabajadoras y sofisticar mecanismos de selección o exclusión plantea retos significativos para el Derecho del Trabajo en la era digital, obligando a repensar las garantías tradicionales frente al poder de dirección empresarial y a reforzar los instrumentos de supervisión, transparencia y rendición de cuentas con perspectiva de género. Solo si se reconoce este poder y se somete a los límites propios de un Estado social y democrático de derecho⁸ será posible evitar que la inteligencia artificial se convierta en un nuevo vehículo de reproducción de las desigualdades entre mujeres y hombres en el trabajo.

⁸ CASAS BAAMONDE M.E. “Tecnologías, Ciencia, Economía y Derecho”, en Derecho y Tecnologías (Dirs) Casas Baamonde y Pérez del Prado, Editorial Universitaria Ramón Areces, 2025, p. 38.

2. MARCO JURÍDICO DE LA IA LABORAL: DATOS SENSIBLES, SUPERVISIÓN Y LÍMITES DEL SECRETO EMPRESARIAL

La integración de las herramientas tecnológicas en la gestión y dirección empresarial sobre las personas trabajadoras supone un auténtico cambio de paradigma en las funciones de los departamentos de recursos humanos y, en general, en el contenido mismo de las relaciones laborales. Desde el inicio de la relación de trabajo, las herramientas algorítmicas se insertan en el poder de dirección empresarial y tienden a convertirse en un instrumento básico de ordenación, control y gestión de los poderes empresariales, desplazando decisiones tradicionalmente humanas hacia sistemas automatizados de evaluación, clasificación y predicción.

Este proceso se produce en un marco normativo en el que el tratamiento de determinados datos personales se encuentra especialmente protegido. El artículo 9 del Reglamento General de Protección de Datos (RGPD) configura como categorías especiales de datos personales –sometidas a una prohibición general de tratamiento salvo concurrencia de base legal específica–, entre otros, los datos que revelen el origen étnico o racial, las opiniones políticas, la filiación sindical, así como los datos relativos a la salud y los referidos a la vida sexual u orientación sexual. El apartado segundo de dicho precepto recoge las excepciones a esa prohibición, incluyendo, en lo que aquí interesa, el tratamiento necesario para el cumplimiento de obligaciones y el ejercicio de derechos específicos del responsable del tratamiento o del interesado en el ámbito del Derecho laboral, siempre que esté autorizado por el Derecho de la Unión o de los Estados miembros.

El RGPD incorpora además dos nociones clave para entender el impacto de la inteligencia artificial en el trabajo: la elaboración de perfiles y las decisiones individuales automatizadas. El artículo 4 define la elaboración de perfiles como cualquier forma de tratamiento automatizado de datos personales destinado a evaluar determinados aspectos personales, tales como el rendimiento laboral, la situación económica, la salud, las preferencias o el comportamiento. El artículo 22, por su parte, reconoce a las personas el derecho a no ser objeto de decisiones basadas únicamente en tratamientos automatizados⁹, incluida la elaboración de perfiles, que produzcan efectos jurídicos sobre ellas o les afecten significativamente de modo similar, admitiendo solo excepciones tasadas y siempre acompañadas de salvaguardas adecuadas para los derechos y libertades del interesado.

La elaboración de perfiles adquiere una relevancia singular en el ámbito laboral porque permite que decisiones esenciales –como la selección de un candidato, la renovación de un contrato temporal o la asignación de incentivos– se adopten a partir de modelos que evalúan o predicen el rendimiento, la “idoneidad” o el comportamiento futuro de las personas trabajadoras. Las decisiones individuales automatizadas, en este contexto, son aquellas que se adoptan sin intervención humana significativa y que producen efectos jurídicos o incidencias relevantes en la esfera profesional de la persona, por ejemplo la aceptación o rechazo de una candidatura en un proceso de selección completamente gestionado mediante algoritmos. El artículo 22 del RGPD erige así un valladar frente a la automatización plena de este tipo de decisiones, pero al mismo tiempo legitima su uso cuando concurren determinadas bases jurídicas y se garantizan salvaguardas como el derecho a la intervención humana, a expresar el propio punto de vista y a impugnar la decisión.

⁹ AGOTE, R. y ARRIOLA, A., “Claves del régimen jurídico de los sistemas de inteligencia artificial para la gestión de las relaciones laborales en España”, *IUSLabor* núm. 3 (2024), p. 159.

En el plano interno, España ha reforzado este marco a través de la Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales. Esta norma modificó el texto refundido de la Ley del Estatuto de los Trabajadores introduciendo el artículo 20 bis, que reconoce expresamente los derechos de las personas trabajadoras a la intimidad en relación con el entorno digital y a la desconexión, y sistematizó los denominados “derechos digitales” en los artículos 87 a 90. De este modo, el ordenamiento laboral incorpora de forma explícita la preocupación por los impactos de las tecnologías de control y gestión, abriendo la puerta a una lectura que conecte el uso de algoritmos con los derechos fundamentales¹⁰ en la empresa, entre ellos la igualdad y la no discriminación.

Sobre este sustrato se superpone ahora un marco regulatorio específico de la inteligencia artificial en la Unión Europea, en plena fase de despliegue tras la entrada en vigor del Reglamento Europeo de Inteligencia Artificial (AI Act). El artículo 6.2 del Reglamento, en relación con el anexo III, clasifica como sistemas de IA de alto riesgo aquellos que afectan al empleo, la gestión de las personas trabajadoras y el acceso al autoempleo, incluyendo de forma expresa los sistemas destinados a ser utilizados para la contratación o selección de personas físicas, así como los que se emplean para adoptar decisiones que influyen en la iniciación, promoción y resolución de relaciones contractuales laborales. Esta calificación como “alto riesgo” no es meramente descriptiva, sino que activa un conjunto reforzado de obligaciones destinadas a asegurar que el uso de la IA sea compatible con la protección de los derechos fundamentales.

Entre esas obligaciones se encuentran, en primer lugar, la implantación de un sistema de gestión de riesgos (artículo 9), concebido como un proceso continuo de identificación, análisis, evaluación y mitigación de los riesgos para la salud, la seguridad y los dere-

¹⁰ MERCADER UGUINA, J.: “Inteligencia Artificial y relaciones laborales”, *Cuadernos de Derecho Transnacional* Vol 16 núm.2 (2024), pp. 1124-1126.

chos fundamentales que pueden derivarse del uso del sistema de IA. En segundo lugar, el Reglamento impone estrictas exigencias de gobernanza de datos (artículo 10), orientadas a garantizar la calidad, pertinencia, representatividad y ausencia de sesgos indebidos en los datos de entrenamiento, validación y prueba, aspecto especialmente sensible cuando se considera la posible reproducción de desigualdades de género en el empleo. A ello se añaden obligaciones de documentación técnica, transparencia y trazabilidad que permitan a los responsables del despliegue interpretar correctamente la información que proporciona el sistema y responder por su comportamiento.

Particularmente relevante resulta el principio de “humano al mando” (humanincommand), que exige que el diseño y desarrollo de los sistemas de alto riesgo permitan una supervisión e intervención humana efectiva (artículo 14). En el ámbito laboral, esta exigencia se traduce en que las personas, y no los algoritmos, deban mantener la última palabra sobre decisiones que afectan de manera intensa a los derechos de las personas trabajadoras, y que dispongan de la información y formación necesarias para ejercer una supervisión crítica. En coherencia con ello, el Reglamento consagra la alfabetización en materia de inteligencia artificial como principio general (artículo 4), imponiendo a proveedores y responsables la obligación de adoptar medidas para que el personal implicado en el diseño, despliegue y uso de la IA alcance un nivel adecuado de conocimientos técnicos y jurídicos en función de su rol.

Desde la perspectiva del Derecho del Trabajo, estas garantías refuerzan la exigencia de transparencia y de negociación cuando la inteligencia artificial impacta sobre las condiciones de trabajo o los criterios de selección. Se apoyan, además, en principios ya asentados en la normativa laboral española, como la buena fe, el deber de información y consulta, o las obligaciones de comunicación de criterios de designación en los despidos colectivos (artículo 51.2 del Estatuto de los Trabajadores) y en los procedimien-

tos de suspensión de contratos y reducción de jornada regulados en el Real Decreto 1483/2012 (artículo 17.2). La combinación de gestión de riesgos, gobernanza de datos, transparencia técnica y supervisión humana configura así un ecosistema de garantías imprescindible para que los sistemas de IA de alto riesgo respeten y hagan efectivos los derechos fundamentales en el contexto laboral, en particular el derecho a la igualdad y a la no discriminación.

En el marco del AI Act, la Unión Europea ha creado también una arquitectura institucional específica para apoyar la aplicación coherente del Reglamento. La Comisión Europea ha establecido un grupo de expertos científicos independientes en inteligencia artificial, cuya función principal es prestar apoyo técnico y científico en relación con los sistemas de alto riesgo y los modelos de propósito general, en estrecha colaboración con la Oficina Europea de Inteligencia Artificial, encargada de coordinar la aplicación del Reglamento en toda la Unión. Este grupo resulta especialmente relevante para el objeto de este estudio porque aporta independencia científica en un ámbito fuertemente tecnificado y contribuye a que la gobernanza de la IA se alinee con los valores fundamentales de la Unión, entre ellos la igualdad de trato entre mujeres y hombres.

El Reglamento de ejecución (UE) 2025/454 de la Comisión detalla la composición y el funcionamiento de dicho grupo de expertos, fijando criterios de selección como el equilibrio de género y la representación geográfica equitativa, así como la independencia respecto de proveedores de sistemas y modelos de IA de uso general. Asimismo, regula la duración de los mandatos, los procedimientos de realización de tareas y la elaboración de dictámenes y documentos técnicos. Esta configuración apunta a dotar a la Unión de una capacidad pericial estable y especializada que actúe como apoyo a las autoridades nacionales, particularmente allí donde los recursos técnicos son limitados y la complejidad de los modelos dificulta su auditoría.

El artículo 14 del AI Act atribuye a las autoridades nacionales de vigilancia del mercado la facultad de solicitar asistencia especializada al grupo de expertos para evaluar sistemas de IA que operan en su jurisdicción cuando existan sospechas justificadas de riesgos significativos, entre ellos la posible discriminación por razón de género, origen étnico, discapacidad u otras categorías protegidas. A través de esta vía, los expertos pueden aplicar metodologías científicas avanzadas para identificar sesgos, proxies discriminatorios o impactos desproporcionados sobre grupos protegidos, reforzando las capacidades institucionales de los Estados miembros en la supervisión de modelos complejos. El propio artículo exige que la solicitud esté motivada, imponiendo a las autoridades la obligación de justificar la existencia de un riesgo real y de acreditar la proporcionalidad de la intervención frente a otros medios disponibles, lo que permite priorizar los casos en los que se aprecian indicios de vulneración del derecho a la igualdad de trato y no discriminación, por ejemplo en forma de sesgos estructurales de género o de impactos sistemáticamente desfavorables sobre las mujeres.

Dado que el artículo 14 permite que las autoridades nacionales recaben apoyo en cualquier ámbito cubierto por el Reglamento, su potencial de actuación se extiende a aquellos sectores donde el riesgo de discriminación estructural es más elevado. El capítulo III y los anexos I y III del AI Act enumeran diversas categorías de sistemas de alto riesgo, entre las que se incluyen, además de los sistemas de cribado y selección de personal, aplicaciones médicas, educativas o vinculadas a la administración de justicia y la seguridad. Si en estos ámbitos se detectan problemas relevantes, las autoridades pueden recurrir al procedimiento del artículo 14 para obtener evaluaciones o peritajes técnicos, favoreciendo una reacción temprana frente a vulneraciones de derechos fundamentales.

Ahora bien, el artículo 14 se limita a habilitar la asistencia técnica, sin enumerar de forma exhaustiva las prácticas concretas que

pueden desplegar los expertos. En el caso español, la Guía de la Agencia Española de Protección de Datos¹¹ sobre “Requisitos para auditorías de tratamientos que incluyan IA” ofrece un referente útil para concretar esas prácticas, al identificar un conjunto de tareas técnicas habituales en auditorías de sistemas algorítmicos. Entre ellas destaca la obligación de incorporar evaluaciones de impacto que incluyan, además de la identificación de riesgos para los derechos y libertades de las personas físicas, la previsión de medidas específicas para afrontarlos, como la minimización de datos, la revisión de la lógica de decisión o la implementación de salvaguardas adicionales.

La guía insiste, asimismo, en la necesidad de que las auditorías no se limiten a fases aisladas del ciclo de vida del sistema, sino que abarquen desde la obtención de los datos y el entrenamiento del modelo hasta su despliegue y mantenimiento, de forma que se identifiquen los puntos críticos en los que pueden ponerse en peligro los derechos de las personas afectadas. Subraya también la importancia de incorporar mecanismos de explicabilidad adecuados al contexto de uso y a los derechos de las personas concernidas, de manera que las técnicas de explicación faciliten tanto la comprensión de las decisiones automatizadas por parte de los responsables del tratamiento como, en su caso, el ejercicio de los derechos de información, oposición o impugnación por las personas trabajadoras.

Otro aspecto central de la guía es la recomendación de incluir comparativas de rendimiento desagregadas por características sensibles –como sexo, edad, origen étnico o discapacidad–, con el fin de medir de forma cuantitativa si el sistema trata de manera distinta a los distintos subgrupos y si existen diferencias relevantes en las tasas de decisión o de error. Para ello, sugiere el uso de métricas de equidad y el registro sistemático de sus resultados como base para la adopción de medidas correctoras, así como el

¹¹ AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS, *Requisitos para auditorías de tratamientos que incluyan IA* (Madrid: AEPD, 2021).

diseño de pruebas específicas para evaluar la variación del comportamiento del sistema entre grupos y contextos. De este modo se refuerza la idea de que la auditoría de la IA debe incorporar metodologías de prueba por grupo y métricas que permitan detectar y corregir sesgos discriminatorios, en particular los que afectan a las mujeres.

La misma guía propone, finalmente, un conjunto de controles a aplicar en las auditorías de tratamientos con IA, con el objetivo de adaptar su alcance al contexto concreto y al nivel de riesgo del tratamiento. Entre estos controles se cuentan la verificación de responsabilidades para asegurar que las decisiones automatizadas están sometidas a supervisión humana efectiva y cumplen la normativa, así como controles específicos sobre los datos, sobre el ciclo de vida del sistema y sobre las pruebas que se realizan para verificar su funcionamiento. Además, se prevé un control de seguimiento y verificación posterior, que permita comprobar que las medidas de mitigación adoptadas se aplican correctamente y mantienen su eficacia a lo largo del tiempo.

En paralelo, el artículo 15 del AI Act establece el procedimiento y los requisitos para que el grupo de expertos científicos pueda solicitar documentación o información a proveedores de modelos de IA de uso general. El artículo 68 concreta las funciones de este grupo, que incluyen la posibilidad de obtener de los proveedores los datos técnicos, operativos o jurídicos necesarios para ejercer sus tareas de evaluación, asesoramiento y supervisión, lo que reviste especial importancia cuando se sospecha que un modelo de propósito general utilizado en contextos laborales pueda estar produciendo efectos discriminatorios.

A estos efectos, el artículo 3.3 del Reglamento define al “proveedor” como toda persona física o jurídica, autoridad pública, agencia u otro organismo que desarrolla un modelo de IA de uso general y lo coloca en el mercado bajo su propio nombre o marca, con independencia de que lo haga a título oneroso o gratuito.

De esta definición resulta que la facultad de requerir información y documentación en virtud del artículo 15 se dirige exclusivamente a los proveedores, quedando excluidos los usuarios que integran el modelo en sus propios sistemas o servicios para uso operativo. En la práctica, esto implica distinguir cuidadosamente entre quien desarrolla y comercializa el modelo y quien lo despliega en un contexto concreto –por ejemplo, una empresa que utiliza una API de IA para la selección de personal–, distinción que es clave para determinar a quién puede dirigirse el grupo de expertos cuando pretenda analizar el funcionamiento técnico del sistema.

Esta delimitación subjetiva genera, sin embargo, importantes complicaciones prácticas. En primer lugar, porque la identificación del proveedor no siempre es sencilla: existen usuarios que modifican, reetiquetan o empaquetan modelos desarrollados por terceros y los colocan en el mercado bajo su propia marca, difuminando la frontera entre proveedor y usuario. En segundo lugar, porque la necesidad de investigar previamente la condición de proveedor de una entidad puede demorar la intervención del grupo de expertos, reduciendo su capacidad de reacción frente a riesgos inminentes. En tercer lugar, porque algunas entidades pueden estructurar deliberadamente sus relaciones contractuales para evitar ser consideradas proveedores y eludir así las obligaciones más intensas del Reglamento. A ello se suma la posible dispersión de la información técnica relevante entre varios proveedores y usuarios, de modo que no siempre resulte claro a quién debe solicitarse la documentación necesaria o sea preciso requerirla a múltiples agentes.

Las consecuencias de estas dificultades son significativas: pueden impedir una evaluación completa de los sistemas de IA de uso general, prolongar las investigaciones y retrasar la adopción de medidas correctoras en presencia de vulneraciones de derechos fundamentales. La necesidad de incoar procedimientos paralelos y coordinar a distintos sujetos incrementa la carga administrativa

para el grupo de personas expertas, la Oficina de IA y la Comisión Europea, y abre la puerta a períodos de impunidad de facto durante los cuales continúan produciéndose decisiones potencialmente discriminatorias sobre personas trabajadoras, en particular sobre las mujeres. No obstante, el propio artículo 15 insiste en la búsqueda de un equilibrio entre el acceso a la información y la protección de la confidencialidad, subrayando que las facultades de obtención de documentación deben ejercerse respetando los secretos comerciales de los proveedores.

En este punto entra en juego el régimen europeo y nacional de protección del secreto empresarial. La Directiva (UE) 2016/943 y su transposición en España mediante la Ley 1/2019, de 20 de febrero, de secretos empresariales, configuran el secreto empresarial como cualquier información que no sea generalmente conocida ni fácilmente accesible, que tenga un valor empresarial y respecto de la cual se hayan adoptado medidas razonables para mantenerla en secreto. La amplitud de esta definición permite que las empresas califiquen como secreto¹² una gama muy variada de informaciones, incluyendo aspectos técnicos del funcionamiento de los sistemas de IA, tales como el código fuente, determinadas arquitecturas de modelo o configuraciones específicas.

De este modo, una empresa proveedora de modelos de IA de uso general sobre la que el grupo de personas expertas desee recabar información puede invocar la protección del secreto empresarial para limitar o condicionar el acceso a documentación clave, siempre que acredite tanto la naturaleza secreta de la información como las medidas de protección adoptadas. En el plano nacional, la amplia comprensión de la noción de secreto empresarial ha generado ya tensiones con otros deberes de información, en particular con el deber de informar a la representación de las personas trabajadoras sobre aspectos relevantes de la organización y

¹² LÓPEZ AHUMADA, J. E., “La exclusión del deber de sigilo de los representantes de los trabajadores de la noción de secretos empresariales: A propósito de la Ley 1/2019 de secretos empresariales, *Temas Laborales* 148 (2019), p. 44.

evolución de la empresa. Ello anticipa conflictos similares cuando se trate de conciliar las facultades de requerimiento del grupo de expertos del AI Act con la protección de los secretos empresariales vinculados a los sistemas de IA.

Nos encontramos así ante un conflicto estructural entre, de un lado, las herramientas que el derecho de la Unión pone a disposición del grupo de expertos para exigir información a los proveedores –siempre que la solicitud sea necesaria y proporcionada– y, de otro, la protección efectiva del secreto empresarial. El artículo 15 obliga a prestar atención a los secretos comerciales, lo que se traduce en que las solicitudes de documentación pueden ser rechazadas o limitadas si no superan el test de necesidad y proporcionalidad, mientras que los proveedores que pretendan reservar determinada información como secreta deberán justificar su carácter y las medidas de protección asociadas. En última instancia, corresponderá a los tribunales de la Unión y de los Estados miembros resolver, caso por caso, este equilibrio entre transparencia y confidencialidad, especialmente cuando estén en juego derechos fundamentales como la igualdad y la no discriminación.

La defensa del derecho de igualdad se ve particularmente comprometida cuando la protección de secretos comerciales se combina con la dificultad de identificar de manera precisa al proveedor responsable del sistema de IA. En escenarios donde intervienen varios actores –distintos proveedores en la cadena de valor y usuarios que integran o afinan el modelo–, puede suceder que ninguno de ellos tenga incentivos para entregar la información que necesita el grupo de expertos o las autoridades nacionales, por temor a revelar elementos de valor competitivo. El riesgo de divulgación inadvertida de secretos empresariales cuando la información se halla distribuida entre varios agentes favorece que cada uno de ellos opte por impugnar las solicitudes, generando procedimientos múltiples y alargando los plazos de investigación.

Pese a estas limitaciones, el artículo 15 encierra un notable potencial como herramienta de protección del derecho a la igualdad y a la no discriminación por razón de género. El principal problema, sin embargo, reside en que la adopción de medidas correctoras frente a sistemas de IA sospechosos de generar desigualdad exige, como condición previa, poder evaluar en profundidad la información técnica del sistema: datos de entrenamiento, métricas utilizadas, registros de decisiones, arquitectura del modelo o configuraciones operativas. Sin acceso a estos elementos, las sospechas de discriminación no pasan de ser conjeturas difíciles de acreditar, lo que vacía de contenido tanto el derecho a la igualdad como las obligaciones preventivas y reparadoras asociadas. Precisamente por ello, el artículo 15 debe entenderse como un instrumento diseñado para diluir la complejidad técnica y hacer frente a la opacidad algorítmica, creando las condiciones para que la discriminación, y en particular la discriminación de género, pueda ser detectada, demostrada y reparada en el marco de un Estado de derecho.

3. DISCRIMINACIÓN ALGORÍTMICA Y “PROBLEMA DEL PROXY” EN EL REGLAMENTO EUROPEO DE INTELIGENCIA ARTIFICIAL

Para situar el Reglamento europeo de inteligencia artificial en el marco del derecho a la igualdad y a la no discriminación es preciso partir del problema específico al que pretende dar respuesta: la denominada discriminación por proxy. La elevada capacidad de inferencia estadística de los sistemas de inteligencia artificial hace posible que, a partir de datos aparentemente neutros y desprovistos de estereotipos, se utilicen variables indirectas que terminan produciendo decisiones discriminatorias¹³. En otros términos, el sistema es capaz de reconstruir, a partir de datos no protegidos¹⁴, información sensible relativa a motivos prohibidos. El núcleo del problema reside en que, para los sistemas de inteligencia artificial, los motivos prohibidos y determinadas características personales tienen un alto valor predictivo, que se explota para anticipar comportamientos humanos¹⁵. Así, los modelos predictivos se entrenan con conjuntos de datos históricos y, cuando estos están sesgados y reflejan desigualdades estructurales, el modelo tiende a reproducirlas y amplificarlas, de modo que los motivos protegidos acumulan efectos de discriminación y pasan a operar como verdaderas señales predictivas.

¹³ KÖCHLING, A. y WEHNER, M. C., “Discriminated by an Algorithm: A Systematic Review of Discrimination and Fairness by Algorithmic Decision-Making in HR.” *Business Research* 13 (2020), pp. 795-848.

¹⁴ HERNÁNDEZ BEJARANO, M., RODRÍGUEZ PIÑERO ROYO, M., TODOLÍ SIGNED, A. y OTROS, *Cambiando la forma de trabajar y de vivir de las plataformas a la economía colaborativa real*, Valencia: Tirant lo Blanch, 2020, pp. 61-64.

¹⁵ RAMÍREZ AUTRÁN, R. “Sesgos y discriminaciones sociales de los algoritmos en inteligencia artificial.” *Entre-textos* 19 (27) (2022): 1–20.

Las características protegidas se correlacionan, por tanto, con resultados observables que los algoritmos detectan, aunque sin reconocer la dimensión normativa que subyace a tales correlaciones. Desde la perspectiva de los proveedores de sistemas de inteligencia artificial, el hecho de prescindir en el entrenamiento de variables explícitas sobre datos sensibles reduce la precisión del modelo, pero no elimina la discriminación. Como se ha demostrado, en ausencia de esas variables el sistema recurre a *proxies* que reproducen, por vías indirectas, el mismo efecto discriminatorio¹⁶.

Además, los sistemas de inteligencia artificial se caracterizan por una capacidad prácticamente ilimitada de procesamiento, que les permite manejar volúmenes ingentes de información en los que inevitablemente se incorporan datos discriminatorios¹⁷. A diferencia de los seres humanos, carecen de una capacidad incorporada para filtrar esos datos a la luz de consideraciones morales o jurídicas. Los algoritmos de aprendizaje automático actualmente disponibles no distinguen entre racionalidad formal y racionalidad sustantiva, lo que les impide interpretar correctamente los valores, normas y juicios morales que acompañan a los datos del mundo social y, en particular, comprender el significado jurídico de la igualdad y la no discriminación¹⁸. A medida que estos algoritmos se perfeccionan, aumentan también su complejidad y su capacidad de identificar y explotar indicadores cada vez más sutiles de motivos prohibidos, de modo que, cuando se les impide utilizar variables protegidas con alta capacidad predictiva, quedan estructuralmente predispuestos a incurrir en

¹⁶ CORBETT-DAVIES, S., GAEBLER, J. D., NILFOROSHAN, H., SHROFF, R. y GOEL, S., "The Measure and Mismeasure of Fairness", *Journal of Machine Learning Research* 24(1) (2023), pp. 37-38.

¹⁷ GE, Y., KNITTEL, C. R., MACKENZIE, D. y ZOEPF, S., Racial and Gender Discrimination in Transportation Network Companies, NBER Working Paper 22776 (2016).

¹⁸ Vid. Prólogo en ARAGÜEZ VALENZUELA, L., *Hacia la eticidad algorítmica en las relaciones laborales*, Laborum, Murcia, 2024.

discriminación por *proxy*¹⁹. La prohibición de utilizar los *proxies* más evidentes no resulta suficiente, ya que la correlación entre los grupos protegidos y otras variables persiste y, si el conjunto de datos es lo bastante rico, el algoritmo puede construir combinaciones indirectas de variables que hacen que la discriminación se vuelva progresivamente más difícil de detectar²⁰.

El Reglamento de Inteligencia Artificial de la Unión Europea reconoce expresamente que la inteligencia artificial tiene usos extraordinariamente beneficiosos, pero también potencial para facilitar prácticas perjudiciales y prohibidas, en particular cuando entra en colisión con los derechos fundamentales consagrados en la Carta, entre ellos el derecho a no sufrir discriminación. El propio Reglamento subraya que, cuando los sistemas de inteligencia artificial no se diseñan o utilizan adecuadamente, pueden vulnerar ese derecho y perpetuar formas de discriminación histórica frente a determinados grupos protegidos²¹. Sin embargo, la discriminación algorítmica no se aborda de manera centralizada en el texto, sino que se menciona de forma específica en el artículo 10.2 f), que remite a la discriminación prohibida por el derecho de la Unión²². Esta remisión sugiere que el legislador europeo asume que la regulación primaria de la discriminación corresponde, ante todo, al acervo ya existente en materia

¹⁹ NISHANT, R., SCHNECKENBERG, D. y RAVISHANKAR, M. N., “The Formal Rationality...”, *Journal of Information Technology* 39(1) (2024), pp. 20-21. <https://doi.org/10.1177/02683962231176842>

²⁰ Prince, Anya and Schwarcz, Daniel, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data* (August 5, 2019). 105 *Iowa Law Review* 1257 (2020), pág. 1257.

²¹ Véase el considerando 28 y 56 del Reglamento de Inteligencia Artificial de la Unión Europea.

²² Establece concretamente este artículo, “...que las prácticas de gobernanza y gestión de datos adecuada para la finalidad prevista del sistema de inteligencia artificial de alto riesgo al que se deberán someter los conjuntos de datos de entrenamiento, validación y prueba deberán centrarse en el examen atendiendo a posibles sesgos que puedan afectar la salud y la seguridad de las personas, aceptar negativamente los derechos fundamentales o dar lugar a algún tipo de discriminación prohibidas por el derecho de la Unión...”

de igualdad y no discriminación, compuesto por las directivas antidiscriminatorias y su desarrollo interno, y que es ese marco normativo el que debe utilizarse para impugnar y reparar la discriminación algorítmica.

Desde esta perspectiva, el Reglamento de Inteligencia Artificial no se configura como una norma laboral, sino como una pieza del derecho de seguridad de los productos de la Unión. Su artículo 2 delimita el ámbito de aplicación atendiendo a los sujetos que participan en el ciclo de vida de los sistemas de inteligencia artificial (proveedores, desplegados, importadores, fabricantes o representantes autorizados), y a esas relaciones dirige fundamentalmente su disciplina. Pese a ello, tampoco puede asimilarse sin más a la normativa clásica sobre fabricación de productos, pues su campo de aplicación material es mucho más amplio y transversal. Aunque el Reglamento no pretende regular directamente los derechos y deberes de las partes en ámbitos sectoriales específicos, como las relaciones laborales, la mejora de la seguridad que persigue tiene un impacto indirecto relevante en la protección de los derechos fundamentales de las personas trabajadoras, al imponer exigencias de diseño y gobernanza que pueden reducir los riesgos de discriminación algorítmica en el empleo²³.

3.1. LA DOCUMENTACIÓN TÉCNICA, LA TRAZABILIDAD Y LA GOBERNANZA DE LOS DATOS EN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL: ALCANCE Y LÍMITE PARA LA PREVENCIÓN DE LOS SESGOS DE GÉNERO

El Reglamento de Inteligencia Artificial impone a los proveedores de sistemas de inteligencia artificial de alto riesgo una obligación reforzada de elaboración y mantenimiento de documentación técnica en el artículo 11, que debe estar disponible antes de la comercialización o puesta en servicio del sistema y mantener-

²³ RODRÍGUEZ CARDO, I. A., “El Reglamento europeo de inteligencia artificial una primera valoración sobre su (limitado) impacto laboral” Trabajo y Empresa núm. 4 (2025), pp. 21-23.

se actualizada durante todo su ciclo de vida, con la finalidad de demostrar el cumplimiento de los requisitos legales y facilitar su evaluación por las autoridades competentes. El contenido mínimo de esta documentación viene definido en el Anexo IV, que enumera ocho²⁴ grandes bloques de información jurídicotécnica y diversos requisitos subordinados, lo que exige una operativización en forma de plantilla estructurada que traduzca esas exigencias normativas en secciones fácilmente auditables.

La primera sección debe contener una descripción general del sistema de inteligencia artificial, especificando, entre otros extremos, la finalidad prevista, los casos de uso contemplados, los límites de uso y la población destinataria, elementos que condicionan la elección de métricas de rendimiento y de equidad, así como la idoneidad de las eventuales medidas de mitigación de sesgos. En esta sección deben documentarse igualmente las modalidades de comercialización y las instrucciones de uso, incluyendo límites operativos y pautas de supervisión humana, a fin de clarificar la distribución de responsabilidades entre proveedor y usuario y de facilitar la imputación de responsabilidad en caso de daños o efectos discriminatorios.

La segunda sección ha de ofrecer una descripción detallada de los elementos del sistema y del proceso de desarrollo, incorporando información sobre la arquitectura del modelo, los algoritmos utilizados, las métricas seleccionadas para evaluar el rendimiento y el impacto discriminatorio –con sus umbrales de aceptabilidad–, así como los mecanismos de intervención humana sobre las decisiones automatizadas. Debe, además, documentar de forma pormenorizada los conjuntos de datos empleados (origen, preprocesamiento, tamaño muestral y representatividad por grupos relevantes), información imprescindible para valorar el riesgo de sesgos –incluidos los de género– o de deriva del modelo a lo largo del tiempo.

²⁴ SOVRANO, F., HINE, E., ANZOLUT, S. y BACCHELLI, A., “Simplifying Software Compliance: AI Technologies in Drafting Technical Documentation for the AI Act”, *Empirical Software Engineering* 30 (2025), Art. 91.

La tercera sección se centra en la monitorización, el funcionamiento y el control del sistema de inteligencia artificial, incluyendo, por un lado, la explicitación de las limitaciones del sistema, entendidas como las condiciones o contextos en los que su rendimiento se degrada de forma significativa, y la identificación de grupos poblacionales para los que la precisión de los resultados puede ser inferior, lo que resulta esencial para el diseño de medidas de mitigación y planes de contingencia. Por otro lado, debe incorporar la documentación de resultados no intencionados – salidas inesperadas, incidentes, medidas correctoras adoptadas y su eficacia, así como escenarios de prueba que revelen comportamientos anómalos –, con el objetivo de detectar fallos emergentes y sesgos que no se manifestaron en las pruebas preliminares y de mejorar la capacidad de detección y respuesta frente a impactos discriminatorios en condiciones reales de uso.

Esta misma sección ha de integrar, además, la identificación y documentación de los riesgos previsibles asociados al funcionamiento del sistema, incluyendo los riesgos de discriminación y su probabilidad e impacto estimados, junto con las medidas técnicas y organizativas de mitigación implantadas y los planes de respuesta a incidentes, en los que se designen claramente los responsables. Esta información es clave para evaluar el equilibrio entre beneficios y posibles daños, así como para que las autoridades puedan valorar la suficiencia de las medidas de mitigación adoptadas; también debe recoger los requisitos relativos a los datos de entrada (estructura, formatos aceptados, criterios de calidad mínima), lo que permite asegurar un funcionamiento estable y reproducible e identificar, en su caso, fuentes de sesgo derivadas de datos deficientes.

La cuarta sección documenta el sistema de gestión de riesgos del artículo 9 del Reglamento, concebido como un proceso iterativo que debe abarcar todo el ciclo de vida del sistema de alto riesgo, con revisiones y actualizaciones periódicas. Esta sección constituye la prueba escrita de que el proveedor no solo identifica los

riesgos –incluidos los sesgos discriminatorios²⁵–, sino que los gestiona activamente, debiendo incluir un inventario de riesgos previsibles (técnicos, jurídicos y operativos), una evaluación de su probabilidad y gravedad, la priorización resultante y las herramientas utilizadas, de forma que pueda comprobarse la proporcionalidad de la gestión de riesgos y que incluso los riesgos residuales queden explícitamente reconocidos.

La quinta sección debe recoger de manera sistemática los cambios introducidos en el sistema desde su diseño hasta su despliegue y explotación actuales, documentando las modificaciones y actualizaciones, así como su impacto en el rendimiento y en la posible generación de nuevos sesgos, mediante evaluaciones y evidencias empíricas. La sexta sección, por su parte, persigue documentar las normas armonizadas, especificaciones técnicas y otros enfoques alternativos empleados por el proveedor para acreditar el cumplimiento de los requisitos legales, describiendo el alcance de cada norma, su ámbito de aplicación dentro del sistema y las limitaciones correspondientes, junto con la evidencia de conformidad (informes de prueba, certificados, resultados de auditorías, documentación de soporte) y el plan de actualización normativa para asegurar la adecuación continua a las futuras evoluciones técnicas y regulatorias²⁶.

La séptima sección debe incluir la declaración formal de conformidad, de acuerdo con los parámetros establecidos en el Anexo V, que actúa como instrumento jurídicotécnico mediante el cual el proveedor acredita que el sistema cumple lo dispuesto en el Reglamento de Inteligencia Artificial, constituyendo prueba oficial ante autoridades y organismos de supervisión. La viabilidad

²⁵ HOLTZ, H. M. y LEDENDAL, J., “AI Data Governance – Overlaps Between the AI Act and the GDPR” Forthcoming in Law, Innovation and Technology, 2025, p. 12.

²⁶ VAN BEKKUM, M. y ZUIDERVEEN BORGESIU, F., “Using Sensitive Data to Prevent Discrimination by Artificial Intelligence: Does the GDPR Need a New Exception?”, Computer Law & Security Review 48 (2023): 105770. <https://doi.org/10.1016/j.clsr.2022.105770>

jurídica y práctica de esta declaración depende, sin embargo, de la solidez de las secciones quinta, sexta y octava, en la medida en que la conformidad no puede entenderse desligada de la documentación sobre cambios, normas aplicadas y seguimiento posterior a la comercialización.

Finalmente, la octava sección tiene por objeto demostrar que el proveedor mantiene un sistema de seguimiento y reevaluación del sistema en condiciones reales de uso, de modo que la conformidad con el Reglamento no sea estática, sino el resultado de un proceso continuo de ajuste. Esta sección debe describir el plan de seguimiento periódico –incluyendo la verificación de que las métricas de rendimiento y de ausencia de sesgos continúan siendo válidas en contextos cambiantes–, los procedimientos de modificación del sistema ante la detección de errores o riesgos, y los mecanismos de información a usuarios y autoridades sobre los problemas detectados, reforzando así la trazabilidad y la rendición de cuentas en materia de discriminación algorítmica²⁷.

Junto a la documentación técnica, el artículo 12 del Reglamento establece obligaciones específicas de registro de eventos para los sistemas de inteligencia artificial de alto riesgo, imponiendo que cuenten con capacidades técnicas de registro automático de acontecimientos durante todo su ciclo de vida, con el fin de asegurar un nivel adecuado de trazabilidad del funcionamiento, facilitar la reconstrucción de operaciones relevantes y apoyar la vigilancia posterior a la comercialización. El precepto exige, como mínimo, el registro de situaciones que puedan convertir el sistema en un riesgo o implicar una modificación sustancial, la conservación de registros que permitan analizar incidentes, fallos y sesgos detectados tras el despliegue, y la documentación necesaria para controlar de forma continuada el comportamiento del sistema; en el

²⁷ VEALE, M., VAN KLEEK, M. y BINNS, R., “Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making”, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, paper 440 (2018), pp. 4-9.

caso de determinados sistemas biométricos de alto riesgo, se añade la obligación de registrar los periodos de uso, la base de datos de referencia, los datos de entrada que generan coincidencias y la identidad de las personas que verifican los resultados, reforzando así la supervisión humana y la asignación individualizada de responsabilidad²⁸.

El artículo 10 se ocupa de la gestión y calidad de los datos utilizados en el desarrollo de sistemas de inteligencia artificial de alto riesgo, exigiendo que los conjuntos de entrenamiento, validación y prueba se sometan a prácticas de gobernanza de datos adecuadas al propósito del sistema y cumplan criterios de calidad tales como relevancia, representatividad y ausencia de errores que puedan inducir sesgos o fallos sistemáticos. De este modo, la norma transforma la gobernanza de datos en un requisito operativo y verificable, al exigir que dichos conjuntos sean auditables y que sus propiedades estadísticas se ajusten al contexto real de utilización del sistema, lo que permite evaluar la conformidad legal y la responsabilidad técnica derivada de los resultados producidos, especialmente en ámbitos sensibles como el empleo y la igualdad de trato.

El apartado 5 del artículo 10 introduce, además, una habilitación específica para que los proveedores de sistemas de alto riesgo puedan tratar categorías especiales de datos personales, siempre que dicho tratamiento resulte necesario para detectar y corregir sesgos generados por el propio sistema, y se respeten plenamente los derechos y libertades fundamentales. Esta previsión compensa en parte las restricciones del artículo 9 del Reglamento General de Protección de Datos, que prohíbe con carácter general el tratamiento de estas categorías salvo excepciones tasadas, y responde a la constatación de que, sin acceso a datos sensibles, muchas auditorías de sesgo –por ejemplo, en relación con el género– resul-

²⁸ SURESH, H. y GUTTAG, J. V., "A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle", arXiv, 28 enero 2019, pp. 4-5.

tan inviables a gran escala, lo que incrementa el riesgo de que las discriminaciones algorítmicas pasen inadvertidas.

El ámbito de aplicación del artículo 10.5 se circunscribe, no obstante, a los proveedores de sistemas de inteligencia artificial de alto riesgo, sin extenderse a usuarios ni operadores del sistema, que permanecen sujetos a las reglas generales de protección de datos y carecen de una base jurídica específica para tratar categorías especiales con fines de auditoría postdespliegue. Esta asimetría genera tensiones prácticas: el proveedor es el único legitimado para procesar datos sensibles con fines de detección de sesgos, pero el usuario es quien dispone de información sobre el funcionamiento contextual del sistema en entornos concretos (por ejemplo, procesos de selección o promoción profesional), lo que dificulta una cooperación efectiva y reduce la eficacia de la excepción como herramienta de prevención y corrección de discriminación en el ámbito laboral.

Además, el origen de los sesgos que desembocan en efectos discriminatorios no se limita a los conjuntos de datos de entrenamiento, validación y prueba en sentido estricto, sino que puede residir también en el contexto sociocultural de aplicación del sistema –marcado por normas y prácticas sociales preexistentes– y en las decisiones de diseño adoptadas por los desarrolladores, desde la definición del constructo objetivo hasta la elección de *proxies* y etiquetas²⁹. Estos elementos contextuales y de diseño pueden provocar que un modelo funcione adecuadamente en una población dominante pero produzca errores sistemáticos en determinados subgrupos, o que incorpore sesgos humanos a través de las etiquetas utilizadas en el entrenamiento, lo que subraya la necesidad de un enfoque holístico de la gobernanza de datos que trascienda el mero control de los conjuntos de datos formales.

²⁹ ALOWAYFI, R., DENNEHY, D., DWIVEDI, Y. y COTTERELL, D., “A Systematic Review of Algorithmic Bias in Decision-Making”, *Proceedings of SaudiCIS 2024*, pp. 2-6.

Finalmente, el artículo 10 adolece de una importante laguna conceptual al no definir qué debe entenderse por “sesgo” ni qué niveles de sesgo resultan aceptables, ni tampoco proporcionar criterios normativos sobre las métricas de equidad apropiadas y las estrategias de mitigación exigibles. Esta indeterminación ha propiciado la proliferación de métricas técnicas de sesgo y equidad desarrolladas en contextos jurídicos diversos, muchas de las cuales se centran en la igualdad formal (igualdad de trato) y no se alinean con el objetivo de igualdad sustantiva perseguido por el Derecho antidiscriminatorio de la Unión Europea, de modo que existe un riesgo real de que se apliquen herramientas de evaluación de sesgo que no se corresponden con los estándares de la jurisprudencia europea en materia de discriminación, incluida la de género.

3.2. CALIDAD DE LOS DATOS, TRATAMIENTO DE CATEGORÍAS SENSIBLES Y SEGOS DE GÉNERO EN EL ART. 10 DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL

Los criterios de calidad de los datos del artículo 10, apartados 2 y 5, pueden agruparse en torno a tres principios básicos: precisión, transparencia y equidad de los datos. Estos principios conectan con los principios éticos no vinculantes recogidos en el considerando 27 del Reglamento de Inteligencia Artificial y en las Directrices Éticas para una IA fiable elaboradas en 2019 por el Grupo de Expertos de Alto Nivel de la Comisión Europea, donde se destacan la transparencia, la diversidad, la no discriminación y la equidad como elementos centrales. A su vez, guardan estrecha relación con los principios del artículo 5 del Reglamento General de Protección de Datos (RGPD), conforme al cual los datos personales deben ser adecuados, exactos, tratados de forma lícita, transparente y para fines determinados, explícitos y legítimos³⁰.

³⁰ OECD, AI, Data Governance and Privacy. Synergies and Areas of International Cooperation, OECD Artificial Intelligence Papers 22 (junio 2024), p. 31.

Sin embargo, la articulación conjunta del RGPD y del Reglamento de Inteligencia Artificial (RIA) genera tensiones normativas relevantes. El artículo 5.1.c) RGPD consagra el principio de minimización, que exige que los datos personales sean “adecuados, pertinentes y limitados a lo necesario” en relación con los fines del tratamiento. Frente a ello, el artículo 10 RIA formula un estándar de exactitud mucho más exigente: los conjuntos de datos de entrenamiento, validación y prueba deben ser pertinentes, representativos, libres de errores y completos según su finalidad, tomando además en consideración el entorno geográfico, contextual, conductual y funcional en el que el sistema va a operar. Esta concepción amplia de la exactitud es la que permite reducir la probabilidad de sesgos sistemáticos –incluidos los sesgos de género–, pero entra en tensión con una lectura estricta del mandato de minimización, que podría desincentivar la recogida de datos relevantes para evaluar desigualdades estructurales³¹.

La Agencia de Derechos Fundamentales de la Unión Europea advirtió ya en 2013, durante la tramitación del RGPD, que era conveniente dejar explícito que la recogida de datos sensibles debía estar permitida cuando fuese necesaria para combatir la discriminación por los motivos enumerados en el artículo 21 de la Carta de Derechos Fundamentales, entre ellos el sexo. No obstante, el artículo 9 RGPD finalmente configuró un régimen muy restrictivo para el tratamiento de categorías especiales de datos, de forma que incluso los tratamientos orientados a salvaguardar derechos fundamentales –por ejemplo, auditorías de sesgo de género– quedan sometidos a condiciones estrictas y a menudo poco operativas en la práctica. El artículo 10.5 RIA pretende responder a esta carencia al prever, de manera excepcional, que los proveedores de sistemas de IA de alto riesgo puedan tratar datos de categorías especiales “en la medida estrictamente necesaria” para detectar y corregir sesgos, pero esta habilitación debe operar

³¹ HOLTZ, H. M. y LEDENDAL, J., “AI Data Governance...”, *Law, Innovation and Technology* (2025), p. 13.

siempre dentro del marco del RGPD, lo que genera incertidumbre interpretativa sobre su alcance efectivo.

En este escenario de aparente conflicto entre la protección de datos y la lucha contra la discriminación, caben dos vías complementarias de solución. En primer lugar, cabe entender el artículo 10.5 RIA como un mecanismo específico de conciliación entre ambos regímenes: el propio precepto vincula el tratamiento de categorías especiales con garantías reforzadas para los derechos fundamentales y lo subordina a la finalidad limitada de supervisar, detectar y corregir sesgos en sistemas de alto riesgo. En segundo lugar, cuando se plantee la tensión entre exactitud y minimización, resulta imprescindible someter el tratamiento de datos sensibles a una evaluación de proporcionalidad, ponderando si la cantidad y el tipo de datos son necesarios y adecuados para cumplir los requisitos de calidad del artículo 10 RIA, al tiempo que se respetan las restricciones del RGPD. Esta evaluación debe incorporar expresamente la dimensión de igualdad y no discriminación, de modo que la ausencia de datos desagregados por sexo u otras categorías protegidas no se convierta, paradójicamente, en un obstáculo para identificar sesgos de género.

El considerando 70 RIA refuerza esta interpretación al exigir que el tratamiento de categorías especiales en virtud del artículo 10.5 se base en una excepción del artículo 9.2 RGPD, señalando de forma particular la excepción por “razones de interés público importante” del artículo 9.2.g). Además, el Tribunal de Justicia de la Unión Europea ha recordado que cualquier tratamiento de datos de categorías especiales debe cumplir, simultáneamente, una base de licitud del artículo 6.1 RGPD, de modo que el artículo 10.5 RIA debe leerse junto con el artículo 6.1 y el artículo 9.2.g), constituyendo en conjunto la base jurídica para el tratamiento de datos personales en este contexto. En términos prácticos, esto supone que la mitigación de sesgos de género en sistemas de alto riesgo solo será jurídicamente viable cuando esté amparada por

una norma de la Unión o de los Estados miembros que configure dicho interés público sustancial y establezca garantías específicas.

El principio de equidad ocupa un lugar central en esta arquitectura normativa. En el RIA se concreta, ante todo, en el artículo 10.2.f), que obliga a examinar los conjuntos de datos de entrenamiento, validación y prueba para detectar posibles sesgos, y en el artículo 10.2.g), que impone la adopción de medidas adecuadas para detectar, prevenir y mitigar esos sesgos, especialmente cuando puedan desembocar en discriminación prohibida por el Derecho de la Unión. En el RGPD, la equidad –artículo 5.1.a)– se entiende como la prohibición de tratar datos personales de forma injustificada, perjudicial, discriminatoria, inesperada o engañosa para las personas afectadas, constituyendo un principio general que se proyecta sobre todo el ciclo de tratamiento³². Ambos preceptos comparten la preocupación por evitar tratamientos discriminatorios, pero difieren en su objeto y en su grado de concreción: mientras el RGPD aplica la equidad al conjunto del tratamiento de datos personales, el artículo 10 RIA la especifica sobre los conjuntos de datos que alimentan sistemas de IA de alto riesgo y traduce el principio en obligaciones operativas muy precisas (auditorías de sesgo, adopción de medidas de corrección, identificación de lagunas de datos).

Desde la perspectiva de género, esta diferencia es crucial. En ausencia de datos desagregados por sexo, identidad de género u otras dimensiones de desigualdad –como la edad, la discapacidad o el origen étnico–, muchos sesgos no se hacen visibles, especialmente los de carácter interseccional. El mandato del artículo 10.2.f) de “examinar los conjuntos de datos en vista de posibles sesgos” obliga, en términos metodológicos, a preguntas específicas sobre la representación de las mujeres y de otros grupos infrarrepresentados, la calidad de los datos relativos a sus trayectorias laborales y la forma en que se codifican las interrupciones de carrera, el

³² EDPB, Guidelines 4/2019 on Article 25 – Data Protection by Design and by Default, versión 2.0 (2020), ap. 69.

trabajo a tiempo parcial o los cuidados no remunerados. Cuando los sistemas se emplean en ámbitos como la contratación, la promoción o la evaluación del rendimiento, la ausencia de controles *ex ante* sobre estas variables puede conducir a la reproducción automatizada de patrones históricos de discriminación de género.

Aun así, la regulación del RIA, por más avanzada que sea en términos de prevención, no garantiza por sí sola la inexistencia de resultados discriminatorios. El enfoque del Reglamento se concentra, sobre todo, en la “entrada” de datos (*data governance*) y en las obligaciones de diseño y desarrollo de sistemas de alto riesgo, pero no agota la protección frente a las “salidas” discriminatorias de los algoritmos. Para estas últimas continúa siendo plenamente aplicable el acervo de Directivas de igualdad de trato de la Unión Europea, así como la jurisprudencia del Tribunal de Justicia en materia de discriminación directa e indirecta por razón de sexo y otras causas, que permiten impugnar decisiones automatizadas que produzcan desventajas particulares para mujeres u otros colectivos. En este sentido, la combinación de normas específicas sobre IA (orientadas a la prevención) y de la tradicional normativa antidiscriminatoria (orientada a la reparación y a la igualdad sustantiva) configura una aproximación dual que puede resultar particularmente fértil en el ámbito de los sesgos algorítmicos.

La excepción del artículo 10.5 RIA, que habilita a los proveedores de sistemas de alto riesgo para tratar datos sensibles con el fin de mitigar sesgos, presenta, con todo, limitaciones importantes: se aplica únicamente a proveedores (no a usuarios ni a otras entidades que podrían detectar sesgos en la fase de uso), se circunscribe al contexto de sistemas de alto riesgo y sujeta el tratamiento a requisitos estrictos de necesidad, proporcionalidad y salvaguardas adicionales. Estas restricciones reducen su utilidad en la fase de despliegue real, en la que muchos sesgos –especialmente los de género, ligados a contextos organizativos concretos– solo se hacen visibles cuando el sistema interactúa con datos reales y con prácticas laborales específicas. Con todo, la excepción no es

irrelevante: proporciona un primer reconocimiento normativo de que, para combatir la discriminación algorítmica, puede ser imprescindible tratar categorías especiales de datos, y constituye un intento inicial del legislador europeo de equilibrar el derecho fundamental a la protección de datos y el derecho a la no discriminación en el contexto de la inteligencia artificial.

En conclusión, los criterios de calidad del artículo 10 RIA, leídos en diálogo con el RGPD y con el Derecho antidiscriminatorio de la Unión Europea³³, ofrecen un marco prometedor pero incompleto para abordar los sesgos de género en sistemas de IA de alto riesgo. La clave residirá en cómo se interpreten y apliquen en la práctica las exigencias de exactitud, equidad y tratamiento de categorías sensibles: si se opta por lecturas formales y minimalistas, se corre el riesgo de invisibilizar desigualdades; si, por el contrario, se adoptan interpretaciones robustas que integren la igualdad sustantiva de género, el potencial de estas normas para transformar la gobernanza algorítmica en un instrumento efectivo de justicia social será considerable.

³³ KAMINSKI, M. E., “Binary Governance: Lessons from the GDPR’s Approach to Algorithmic Accountability”, *Southern California Law Review* 92 (2018).

4. DISCRIMINACIÓN ALGORÍTMICA

4.1. DATOS QUE REFLEJAN DISCRIMINACIÓN ALGORÍTMICA

La inteligencia artificial y su industria está transformando la naturaleza del trabajo³⁴ influyendo gravemente en las dinámicas de género y agravando la inseguridad laboral sobre todo para las mujeres. Uno de los aspectos que más preocupa es la exclusión de las mujeres de la participación en la fuerza laboral digital. La infrarrepresentación de las mujeres en los sectores tecnológicos, lo que actualmente incluye el manejo de la inteligencia artificial, no se debe ningún caso a desinterés o a falta de cualificación.

Según el Instituto de las Mujeres del Gobierno de España el total de mujeres matriculadas, según el ámbito de estudio, informática en el curso 2024 y 2025 suponía un 17,7 % del total, suponiendo un incremento de 0,6 puntos porcentuales con respecto al curso anterior³⁵. Por otra parte, según la Eurostat el porcentaje de todos los graduados de educación terciaria en campos de educación STEM que son mujeres supone el 33,5 % del total para el año 2023, el último año del que se tiene información. Mientras que en España para el mismo año el 27,2% del total de personas graduadas en educación terciaria en este campo eran mujeres, habiéndose producido un incremento de 0,3 puntos porcentuales en comparación con el año 2024, el último año del que se tiene

³⁴ MERCADER UGINA J.: "Algoritmos e inteligencia artificial en el derecho digital del trabajo" en *Derecho y Tecnologías* (dir. Casas Baamonde y Pérez del Prado), Editorial Universitaria Ramón Areces, 2025, p. 335.

³⁵ INSTITUTO DE LA MUJER, "Alumnado universitario – MujerCifras", consultado 14 febrero 2026.

registro de datos, en el que el porcentaje es de 27,5. Por lo tanto, en España el porcentaje de mujeres de educación terciaria en el campo STEM es 6 puntos porcentuales por debajo de la media de la Unión Europea. Además, comparado con los demás países de la Unión Europea España se queda entre los últimos países, únicamente superado por Suiza, en el que solo el 26,3 % del total de graduados de educación terciaria en este campo son mujeres, mientras que la mayoría de países tienen un porcentaje por encima del 30 % siendo el país que mayor porcentaje de mujeres graduadas en este campo tiene que Bosnia-Herzegovina³⁶.

Conforme el informe de datos y cifras, sobre la distribución del personal docente investigador en los centros de universidades públicas, según el grupo de edad y el ámbito de estudio, en el curso 2023 y 2024 en el ámbito de informática, del 4,2 % solo el 1,9 % eran mujeres³⁷. Según el informe científicos en cifras del año 2025, la evolución del porcentaje de alumnas matriculadas en estudios de doctorado en la rama de enseñanza de ingeniería y arquitectura aunque ha crecido desde el curso 2014-2015 hasta el 2023-2024 en estos 10 años solo lo ha hecho en menos de un 1 %, pasando del curso 2014-2015 de ser un 29,4 % mujeres del total de matriculados en estudios de doctorado en ingeniería y arquitectura a ser en el curso 2023-2024 un 31,1 %. Esta infrarrepresentación contrasta con la representación casi igualitaria que tiene las mujeres en otras ramas de investigación y enseñanza, como las artes y humanidades y las ciencias de la salud. Según el mismo informe, las mujeres solo ocupan el 12,6 % del total de la distribución del personal investigador en enseñanza superior en la disciplina de ingeniería y tecnología, mientras que los hombres ocupan el 26 %, siendo la disciplina científica donde tienen más presencia. Si nos centramos en el personal investigador de las diferentes etapas, siendo la primera etapa la que representa a los jóvenes investigadores o investigadoras que

³⁶ EUROSTAT, “Students by field of education (tps00217)”.

³⁷ MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES, Datos y cifras del Sistema Universitario Español. Curso 2025-2026 (Madrid, 2025), p. 174.

empiezan la carrera y la última etapa al profesorado catedrático de universidad. En el curso 2022-2023, del último del que se tiene información, en cuanto a la evolución de la distribución del personal investigador de grado A en las universidades, el área de ingeniería y tecnología es el que presenta mayor brecha de género, con una diferencia de 8,2 puntos porcentuales, siendo la participación de las mujeres del 11 % frente al 19 % de los hombres, brecha de género que se lleva manteniendo desde hace 10 años³⁸.

Conforme al estudio de 2023 sobre la situación de la investigación en informática en España, la conciliación laboral se entiende como una herramienta que repercute de manera positiva tanto a las personas trabajadoras como a los empresarios. Sin embargo, tras analizar los datos de la encuesta se observa que las mujeres tienen más dificultades que los hombres a la hora de cumplir con sus responsabilidades laborales o académicas debido a sus responsabilidades personales. Además, aunque independientemente del sexo las personas que tienen personas a cargo son las que tienen más dificultades, siendo la diferencia tanto en el caso de mujeres como de hombres, de casi 40 puntos porcentuales entre las personas con otras personas a cargo y las que no tienen a nadie a cargo, en particular se aprecia que las mujeres con personas a cargo se distancian de los hombres con personas a cargo por más de 10 puntos porcentuales, concretamente un 62 % comparado con un 76 % de mujeres a las que les representan dificultades las responsabilidades personales. Por otra parte, en relación con el avance de la carrera profesional, un factor importante es la financiación, por lo que la falta de financiación siendo una causa del abandono de la carrera profesional tecnológica es considerada un obstáculo importante por el 52,2 % de mujeres, superando la cifra de los hombres por más de cinco puntos porcentuales, pero lo importante de este factor es que el porcentaje de mujeres que considera que la falta de financiación que es un obstáculo importante

³⁸ MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES, Científicas en cifras 2025 (Madrid, marzo 2025), p. 50.

aumenta en el caso de mujeres con personas a cargo elevándose a un 56% en comparación con un 45% de hombres con personas a cargo. Centrándonos en cómo han percibido la discriminación por sexo en los encuestados, lo que llaman la atención es que un 69.1% de las mujeres indica que creen que la discriminación por sexo afecta más a mujeres que hombres³⁹.

Por último, conforme al gráfico del informe de mujeres en STEM sobre la evolución del porcentaje de mujeres y hombres con ocupaciones en estas áreas, se puede ver una línea ascendente desde el 2011 al 2022 del porcentaje de mujeres, pasando de suponer un 3,3% en el total de ocupaciones a un 5,5 % en 2022, observándose sin embargo una brecha de género de al menos 5 puntos porcentuales en todos los años. Por otra parte, en cuanto a la proporción de hombres y mujeres en las ocupaciones en STEM, las mujeres solamente representan el 25% de los técnicos STEM, frente al 75% de técnicos hombres, asimismo solo representan el 27% en las ocupaciones catalogadas como profesionales en STEM. Esta enorme brecha de género se puede ver desde el 2011 hasta el 2022, donde el porcentaje de mujeres alcanzó como máximo el 28% en el 2018. Además, de las personas graduadas en carreras STEM solo el 39,8% de mujeres terminan siendo después de cinco años profesionales STEM mientras que el 46,2% de hombres llega a serlo, siendo la diferencia de alrededor de siete puntos porcentuales. Por otra parte, también existe una brecha salarial en los salarios promedio de mujeres y hombres, concretamente en 2022 el salario de los profesionales STEM hombres multiplicaba por 1,10 el de las mujeres. La brecha de género también es visible en el porcentaje de hombres y mujeres con contratos a tiempo parcial, en el que el porcentaje de mujeres con contrato a tiempo parcial multiplica por 2,51 el de los hombres⁴⁰.

³⁹ GONZÁLEZ GONZÁLEZ, C. S., GARCÍA, I. y OTROS, Situación de la investigación en informática en España: Informe Women@Inf (SCIE y Universidad de La Laguna, 2023).

⁴⁰ COBREROS, L., GALINDO, J., RAIGADA, T. y ISLA, C., *Mujeres en STEM: Desde la educación básica hasta la carrera laboral*. Barcelona: EsadeEcPol – Center for Economic Policy, 2024.

La falta de trabajos flexibles en STEM, unido a que las mujeres suelen necesitar más flexibilidad para la conciliación puede explicar porque muchas no eligen o abandonan estas trayectorias. Conforme a un estudio, la maternidad se convierte en un factor determinante en la salida de las mujeres del mercado laboral en la rama de ingeniería, por tanto las ideas sobre retención de género deben incorporar la interacción entre el ciclo vital reproductivo y la carrera profesional. Por otra parte, la diferencia entre rangos de edad no se explica únicamente por los cambios en edad de casarse, sino que existen efectos específicos por cada subgrupo de mujeres. Además, la situación familiar incide en la relación de la variable de género y la de continuidad laboral, de esta forma se entiende que las mujeres que no tienen cargas familiares tendrán una trayectoria mayor que las mujeres con hijos. Teniendo en cuenta que la salida de los hombres tiene una variación muy baja, se entiende que el análisis de las dinámicas de género tiene que centrarse en los factores que afectan diferencialmente a las mujeres, por ejemplo en las políticas laborales. Además, en los estudios para la mayor retención de las mujeres en la trayectoria de esta ocupación, debe tener en cuenta en todo momento las interacciones que se producen entre las variables de género, situación familiar y grupos de edad ya que en caso de omitirlas, se estarían subestimando diferencias críticas. En cuanto a la maternidad, las mujeres con hijos tienen muchas menos probabilidades de permanecer en ingeniería llevando entre siete y ocho años de carrera en todos los rangos de edad analizados que los hombres, sin embargo, incluso el matrimonio sin hijos también reduce la probabilidad de que las mujeres continúen en su carrera en ingeniería en algunos rangos de edad. Por tanto, se concluye que en la interacción entre la variable género y la situación familiar todas las mujeres, independientemente de la situación familiar que tengan, tienen más probabilidades que los hombres de abandonar su ocupación en ingeniería, pero además, entre las mujeres que continúan trabajando a tiempo completo, aquellas mujeres que tienen hijos tienen aún mayores probabilidades de abandonar la

ocupación de ingeniería, sobre todo en los contextos en los que no existe posibilidad de tener flexibilidad horaria o se exigen largas jornadas⁴¹.

Conforme el estudio de la Unión Europea, a nivel comunitario incluyendo todos los campos académicos, a mayor avance en el mundo académico para las mujeres, menor es su representación, sobre todo en los puestos de más categoría o nivel. Las mujeres siguen estando infrarrepresentadas en los puestos de grado mayor en todos los campos, representando solo el 30 % de los puestos de mayor grado del total. Está su representación supera en todos los casos el 30 %, sin embargo, para los campos de ciencia e ingeniería la brecha de género es mayor, ya que las mujeres solo ocupan el 20 % de los puestos de este tipo⁴². En España, conforme a la herramienta Comunitaria denominada “She Figures Index” se ha podido medir que la cantidad de mujeres que ocupaban puestos de grado viene incrementándose desde 2013 hasta 2022, habiendo pasado del 21 % al 24 %, sin embargo España continúa estando por debajo de la media de la Unión Europea en cuanto a la proporción de estos puestos. Siendo más concretos, en cuanto al acceso de las mujeres a puestos de liderazgo y tomas de decisiones, desde 2014 a 2022 se ha producido un aumento de casi 10 puntos porcentuales, situándonos en 2022 con una representación de mujeres del 44 % del total de miembros de los consejos⁴³.

En 2018, ya existían resultados que indicaban que conforme el avance de la tecnología, el 10 % de la fuerza laboral tanto masculina como femenina en 30 países, contando a los miembros de la OCDE estaban en riesgo alto, lo que significaba que existía una

⁴¹ KAHN, S. y GINTHER, D. K., “Are Recent Cohorts of Women with Engineering Bachelors Less Likely to Stay in Engineering?” *Frontiers in Psychology* 6 (2015): 1144, p. 12.

⁴² EUROPEAN COMMISSION, *She Figures 2024: Gender Equality in Research and Innovation* (Brussels, 2024).

⁴³ EUROPEAN COMMISSION, *She Figures 2024 – Spain Country Fiche* (Brussels, 2024).

probabilidad mayor al 70 % de que sus puestos de trabajo fuesen automatizados y las próximas dos décadas, dentro de estos datos la fuerza laboral femenina tenía una proporción mayor en alto riesgo de automatización, concretamente, el 11 % de la fuerza laboral femenina estaba en alto riesgo, frente a un 9 % de fuerza laboral masculina. Sobre todo, aquellos supuestos en los que se veía una mayor probabilidad de automatización eran en los casos donde al factor femenino se le sumaban otros factores como bajo nivel educativo, puestos administrativos y una edad mayor a 40 años. Por lo tanto, las brechas de género en la amenaza de automatización tienen como una de las principales causas las diferencias en las características de los puestos de trabajo. Por ejemplo, el sector servicios es uno de los que enfrenta un mayor riesgo de automatización, sin embargo en el mismo las mujeres están sobrerrepresentadas y enfrentan un mayor riesgo que los hombres con sus mismos puestos de trabajo, sobre todo porque existe más probabilidades de que sean los hombres quienes tengan tareas más abstractas y por tanto más difíciles de automatizar. Por otra parte, en los supuestos de mayor nivel existe una mayor probabilidad de que las mujeres se enfrentan a un mayor riesgo de automatización que los hombres⁴⁴.

Los casos de sesgo en el aprendizaje automático que surgen cuando un algoritmo produce resultados sesgados debido a supuestos erróneos se deben principalmente a que los algoritmos de aprendizaje automático se basan en datos del mundo real por lo que contienen los sesgos presentes en la sociedad. El sesgo de género es uno de estos posibles supuestos erróneos, así también sucede con los sesgos de raza o religión. Para intentar corregir estos sesgos, unos factores muy importantes son la transparencia y la explicabilidad, que analizaremos más adelante, pero relacionado con la brecha de género en los puestos de liderazgo, la diversidad en los equipos que desarrollan los sistemas de inteligencia artifi-

⁴⁴ BRUSSEVICH, M., DABLA-NORRIS, E., KAMUNGE, C., KARNANE, P., KHALID, S. y KOCHHAR, K., Gender, Technology, and the Future of Work (IMF) en: www.imf.org/-/media/files/publications/sdn/2018/sdn1807.pdf

cial puede ser un factor crucial para poder identificar situaciones injustas, sírvase de ejemplo investigadoras que son líderes en el campo de la detección y corrección de sesgos en algoritmos inteligentes. Por tanto, es crucial que las mujeres participen activamente en definir lo que se entiende por equidad, además de en las decisiones de métricas, datos y objetivos para los sistemas de inteligencia artificial ya que no solo se trata de una cuestión de justicia, sino que también sirve como una medida para que los avances en igualdad de género continúen⁴⁵.

4.1.1. IA y brecha de género: una radiografía estadística

Los datos disponibles⁴⁶ confirman que la expansión de la inteligencia artificial se produce sobre una base de desigualdad de género ya muy marcada, tanto en el acceso a los empleos tecnológicos como en la exposición a los riesgos de automatización.

En primer lugar, la presencia de mujeres en los perfiles profesionales directamente vinculados a datos e inteligencia artificial sigue siendo claramente minoritaria. En España, en 2023 las mujeres representaban en torno al 32 % del conjunto de profesionales de datos e IA, frente al 68 % de hombres, lo que refleja una brecha persistente en un ámbito decisivo para el diseño y el control de los sistemas algorítmicos. A escala global la situación es aún más acusada: distintos estudios estiman que alrededor del **26-29** % de los empleos relacionados con IA están ocupados por mujeres, con porcentajes todavía menores en nichos como la computación en la nube o la ciencia de datos. Esta infrarrepresentación implica que las decisiones sobre qué problemas se automatizan, con qué datos y bajo qué criterios de éxito se toman mayoritariamente desde miradas masculinas, reduciendo la probabilidad de

⁴⁵ Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning." IEEE/ACM Workshop on Gender Equality in Software Engineering (2018): 14–16.

⁴⁶ https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/12/algorithm-and-eve_0e889c45/a1603510-en.pdf

que los sesgos de género sean detectados y corregidos de manera temprana⁴⁷.

En segundo lugar, la brecha de género se manifiesta también en el acceso a las competencias digitales avanzadas necesarias para participar en la economía de la IA. Informes recientes señalan que, a escala mundial, las mujeres tienen menos probabilidades que los hombres de adquirir habilidades digitales básicas y son hasta cuatro veces menos propensas a contar con capacidades avanzadas de programación, lo que limita su presencia en los segmentos mejor remunerados y con mayor poder de decisión. En Europa, aunque las tasas de empleo femenino en ocupaciones expuestas a la IA han aumentado, los análisis muestran que la digitalización tiende a reforzar la segregación horizontal y vertical si no se acompaña de políticas activas de formación y promoción profesional para las mujeres⁴⁸.

Por último, diversos estudios sobre el impacto de la IA en el mercado de trabajo apuntan a un doble riesgo para la igualdad. Por un lado, determinadas ocupaciones altamente feminizadas –como servicios administrativos, atención al cliente, determinados segmentos de servicios sociales o tareas rutinarias en sanidad y educación– presentan un elevado potencial de automatización, lo que podría concentrar sobre las mujeres una parte desproporcionada de la destrucción de empleo. Por otro, la limitada presencia femenina en los puestos de diseño y liderazgo de la IA dificulta que las oportunidades de creación de nuevos empleos de calidad en este ámbito se distribuyan de forma equitativa, alimentando un “círculo vicioso de desigualdad digital” en el que la menor participación de las mujeres en la IA refuerza las brechas de participación laboral, salarios y poder de decisión⁴⁹.

⁴⁷ GÓMEZ-HERRERA, E. y S. KOESZEGI (2022) ‘Una perspectiva de género sobre la inteligencia artificial y el empleo: el círculo vicioso de la desigualdad digital’, *Documento de trabajo* 15/2022, Bruegel, <https://www.bruegel.org/working-paper/gender-perspective-artificial-intelligence-and-jobs-vicious-cycle-digital-inequality>

⁴⁸ <https://www.sheai.co/blog/women-in-ai-key-statistics>

⁴⁹ https://www.imk-boeckler.de/data/downloads/OEA/Veranstaltungen/2025/v_2025_10_25_qanas.pdf

En conjunto, estas estadísticas muestran que la inteligencia artificial se despliega en un terreno de juego profundamente desigual. La combinación de infrarrepresentación femenina en los empleos de IA, déficit de competencias digitales avanzadas y concentración de mujeres en ocupaciones con alto riesgo de automatización configura un escenario en el que la transición digital puede agravar las brechas de género si no se introducen políticas robustas de corrección: desde estrategias de formación y reciclaje profesional específicas para mujeres, hasta medidas de acción positiva en la contratación y promoción de perfiles de IA, pasando por la incorporación obligatoria de indicadores desagregados por sexo en todas las evaluaciones de impacto de la IA en el empleo⁵⁰.

4.2 VULNERACIÓN DE DERECHOS FUNDAMENTALES: DERECHO A LA NO DISCRIMINACIÓN

En la doctrina española sobre tutela de derechos fundamentales en el ámbito laboral, el Tribunal Supremo ha insistido en que la inversión de la carga de la prueba exige, como presupuesto inexcusable, la previa acreditación por la parte demandante de indicios razonables de la lesión alegada⁵¹. No se requiere prueba plena, pero sí una actividad probatoria suficiente, apoyada en hechos concretos y objetivos, de los que pueda desprenderse una presunción o apariencia de discriminación. En la misma línea, el Tribunal Constitucional define la prueba indiciaria como un hecho o un conjunto de hechos acreditados a partir de los cuales resulte posible deducir, con un juicio de probabilidad cualificada, la existencia de una vulneración del derecho fundamental, siempre que tales hechos generen una sospecha razonable, apariencia o presunción de lesión⁵².

⁵⁰ Distribución por género de los profesionales de IA y datos en España en 2023, <https://www.statista.com/statistics/1491506/spain-artificial-intelligence-professionals-gender-distribution>

⁵¹ STS de 21-02-18, rec. 842/2016.

⁵² STC 104/2014, de 23 de junio.

Aun cuando la jurisprudencia constitucional y del Tribunal Supremo no se refiere siempre de forma expresa a la denegación total de acceso a la información por parte de la empresa como elemento autónomo de prueba, esta conducta puede integrarse dentro del haz de hechos que conforman la prueba indiciaria. En este sentido, la negativa injustificada a facilitar información relevante –por ejemplo, datos sobre el funcionamiento de un sistema de decisión automatizada– puede constituir un indicio adicional de discriminación directa o indirecta, en línea con la interpretación del Tribunal de Justicia de la Unión Europea (TJUE), que ha reconocido que tales negativas pueden ser valoradas como elementos probatorios en el marco del derecho antidiscriminatorio europeo.

Cuando se discute la existencia de discriminación en el ámbito laboral, el cauce procesal específico es el procedimiento de tutela de derechos fundamentales y libertades públicas regulado en los artículos 177 y siguientes de la Ley reguladora de la jurisdicción social (LRJS). Siempre que en un proceso se alegue la vulneración de un derecho fundamental –incluida la prohibición de trato discriminatorio– resultan aplicables las reglas específicas de este proceso, en particular las relativas a la carga de la prueba. Conforme al artículo 181 LRJS y a la doctrina consolidada del Tribunal Constitucional, una vez acreditados indicios de trato discriminatorio o lesivo de derechos fundamentales, la carga de la prueba se desplaza al empresario, que debe aportar una justificación suficiente, razonable y ajena a todo móvil ilícito que explique la medida adoptada, incluidas aquellas decisiones adoptadas mediante o con apoyo en sistemas de inteligencia artificial⁵³.

Esta inversión de la carga probatoria no exime a ninguna de las partes de su deber de colaboración probatoria: la persona demandante ha de acreditar indicios racionales de lesión, mientras que la parte demandada debe demostrar la razonabilidad y proporcio-

⁵³ STSJ Cataluña, 6092/2025, de 17 de noviembre de 2025 (rec. 3063/2025).

nalidad de su actuación. El Tribunal Constitucional ha subrayado que solo cuando se hayan probado indicios consistentes de vulneración puede imponerse al empleador la carga de acreditar que su decisión responde a causas objetivas y razonables, totalmente ajenas a cualquier propósito o efecto discriminatorio. De ello se desprende que, aun existiendo inversión de la carga de la prueba, ambas partes mantienen una actividad probatoria, si bien de distinta intensidad y con exigencias diferenciadas.

En presencia de indicios de discriminación, corresponde a la empresa ofrecer una justificación objetiva y razonable, suficientemente acreditada, de las medidas empresariales controvertidas y de su proporcionalidad respecto del fin perseguido. La parte demandada puede optar, esencialmente, por dos estrategias: negar la existencia de vulneración –discutiendo el vínculo causal entre un motivo protegido y el resultado diferenciado obtenido, por ejemplo, a través de un sistema de IA– o bien admitir la existencia de un impacto diferenciado pero sostener que está amparado por una causa legítima y proporcionada, ajena a cualquier finalidad discriminatoria. La jurisprudencia constitucional ha ido consolidando la idea de que la inexistencia de intención discriminatoria no exime al empresario de responsabilidad si el resultado objetivo de su actuación implica una vulneración del derecho fundamental a la igualdad y a la no discriminación⁵⁴; lo relevante es el efecto producido, no el propósito subjetivo.

El Tribunal Supremo ha precisado que el artículo 181.2 LRJS no comporta una “transformación pura” del régimen general de la carga de la prueba, sino una modulación que produce efectos concretos sobre la posición procesal de las partes. En particular, se exige a la persona trabajadora únicamente la aportación de indicios racionales de la lesión invocada, lo que, en teoría, rebaja el umbral probatorio respecto de la prueba directa del móvil discriminatorio empresarial. Sin embargo, en la práctica, la parte

⁵⁴ STC de España de 21 de marzo de 1986.

trabajadora continúa en una posición de desventaja estructural, especialmente cuando la discriminación deriva, total o parcialmente, del funcionamiento de sistemas de inteligencia artificial, respecto de los cuales el acceso a la información técnica, a los datos de entrenamiento o a los criterios de decisión suele estar en manos exclusivas del empleador o del proveedor tecnológico.

Esta asimetría probatoria se agrava si se tiene en cuenta que, en el plano del Derecho de la Unión, la evidencia estadística desempeña un papel particularmente relevante para acreditar o refutar la existencia de un vínculo causal entre los motivos protegidos y los resultados diferenciales. La práctica demuestra que buena parte de estas pruebas estadísticas son aportadas por las empresas demandadas, que utilizan como referencia al grupo presumiblemente favorecido para intentar desvirtuar la alegación de exclusión o desventaja de determinados colectivos. Aunque, en abstracto, ambas partes pueden presentar pruebas estadísticas, en la realidad la disponibilidad de datos desagregados por sexo, edad u otras categorías protegidas suele concentrarse en el empleador —y, en su caso, en el proveedor del modelo de IA—, lo que dificulta seriamente que la persona trabajadora pueda construir una prueba indiciaria robusta, especialmente en supuestos de discriminación indirecta y de sesgos algorítmicos de género.

De cara al futuro, todo apunta a que la relevancia de la prueba estadística en litigios de discriminación laboral⁵⁵ aumentará en paralelo al uso intensivo de sistemas automatizados de decisión y al potencial de discriminación indirecta que éstos incorporan⁵⁶. Si bien el avance del aprendizaje automático y el riesgo de sesgos no puede detenerse, sí es posible reforzar las exigencias de calidad probatoria y de razonamiento jurídico que se imponen a las empresas. En esta línea, resulta significativo que el TJUE haya ido endureciendo el escrutinio de las pruebas estadísticas presentadas

⁵⁵ STC 51/2021; STC 31/2014; STC 17/2007; STC 173/2013.

⁵⁶ STS de 5 de diciembre de 2000, rec. 4374/1999; STSJ de Madrid 672/2025, 3 de noviembre de 2025(rec. 350/2025).

por las partes, insistiendo en la necesidad de analizar de forma coherente tanto el grupo presuntamente perjudicado como el favorecido y de tomar en consideración los patrones de desigualdad estructural que afloran en la comparación. Un incremento de las exigencias probatorias para la parte demandada no sustituye la necesidad de reforzar la posición procesal de las personas trabajadoras, pero sí contribuye a corregir parcialmente la desigualdad de armas probatorias que se acentúa cuando la discriminación se vehicula mediante sistemas de inteligencia artificial.

4.3 OPACIDAD ALGORÍTMICA, TRAZABILIDAD DEL "PIPELINE" Y RIESGOS DE DISCRIMINACIÓN

Un sistema de inteligencia artificial puede ser opaco en la medida en que no resulta posible comprender ni reconstruir el modo en que transforma sus entradas en salidas, incluso cuando se conoce su arquitectura técnica. La opacidad tiene múltiples fuentes. Una primera fuente es la fuerte dependencia contextual del comportamiento del modelo: aunque el diseño sea conocido, la respuesta efectiva del sistema varía según el entorno organizativo, los datos disponibles y las interacciones con otros sistemas, lo que dificulta anticipar sus resultados en situaciones concretas. Una segunda fuente procede de la propia complejidad técnica de determinados modelos, especialmente los basados en aprendizaje profundo, caracterizados por arquitecturas de redes neuronales con funciones no lineales y millones de parámetros, de modo que, aun disponiendo del código, resulta prácticamente imposible reconstruir, en términos comprensibles para personas no expertas, cómo se llega de determinadas variables de entrada a una decisión u otra.

Una tercera fuente remite a la insuficiencia de información en la documentación técnica o en los artefactos que acompañan el ciclo de vida del sistema (datasets, modelos entrenados, versiones, logs, etc.). La falta de registros sistemáticos, de criterios claros de

gobernanza de datos o de alfabetización técnica dentro de la organización hace que no pueda comprenderse adecuadamente cómo el sistema produce sus resultados. Esta opacidad puede ser involuntaria –derivada de una mala gobernanza o de la ausencia de cultura técnica–, pero también deliberada, cuando proveedores o usuarios califican como secreto empresarial tanto el código como los datos, someten la información a acuerdos de confidencialidad o la ocultan estratégicamente para preservar ventajas competitivas, dificultando cualquier auditoría independiente⁵⁷.

Además de las causas, la literatura distingue dos grandes tipos de opacidad en función del ámbito en el que se proyecta su impacto: opacidad del mecanismo y opacidad de las entradas. La primera se produce cuando no se puede entender cómo decide el algoritmo, es decir, cómo transforma las entradas en salidas; la segunda aparece cuando se desconoce qué datos utiliza el sistema, cómo se seleccionan, cómo se preprocesan o qué variables derivadas se construyen antes de alimentar el modelo. Estas dos formas suelen interactuar: los modelos muy complejos tienden a generar principalmente opacidad del mecanismo, porque aun con acceso al código resulta imposible extraer una explicación inteligible de la combinación de variables que lleva a una determinada salida; la falta de documentación y registros, por su parte, produce sobre todo opacidad de las entradas, al impedir reconstruir qué información concreta se utilizó, cómo se limpió o qué transformaciones sufrió⁵⁸.

No obstante, ciertas causas pueden producir simultáneamente ambos tipos de opacidad. Los secretos comerciales y los acuerdos de confidencialidad bloquean tanto el acceso al proceso interno del sistema como el conocimiento del origen y características de los datos, erigiendo un doble muro frente al control externo. La baja alfabetización técnica de quienes gestionan o supervisan el

⁵⁷ <https://aesia.digital.gob.es/en/presentria-sistemas-de-alto-riesgo>

⁵⁸ Art. 10 Data and data governance, AI Act Explorer.

sistema implica que, aun existiendo documentación suficiente y acceso al código, estos recursos no se traducen en una reducción efectiva de la opacidad, porque no se dispone de competencias para interpretarlos de manera crítica. Incluso factores organizativos como la alta rotación de personal en las empresas proveedoras pueden provocar pérdida de conocimiento tácito, desaparición de “memoria institucional” y ausencia de registros adecuados, lo que se materializa en más opacidad de las entradas y del propio mecanismo de decisión.

Desde la perspectiva de la igualdad y la no discriminación, el impacto de la opacidad es especialmente grave. Si solo existe opacidad del mecanismo, resulta difícil entender por qué el sistema trata de manera desigual a distintos grupos; si además hay opacidad de las entradas, la situación es todavía más problemática, porque no se sabe qué datos sustentan los patrones discriminatorios, lo que impide diseñar estrategias de mitigación, identificar las variables con efecto dispar y, en última instancia, demostrar jurídicamente la existencia del trato desigual. La ausencia simultánea de visibilidad sobre “cómo decide” y sobre “con qué datos decide” configura un escenario de máxima vulnerabilidad para las personas afectadas, que se enfrentan a decisiones potencialmente lesivas sin posibilidad real de escrutinio ni de impugnación informada⁵⁹.

La capacidad predictiva de un sistema de inteligencia artificial depende, en gran medida, de la calidad y diversidad de los datos de los que se nutre y de las decisiones de modelización que se adoptan. Cuando los datos históricos incorporan sesgos estructurales –por ejemplo, infrarrepresentación de mujeres en ciertos puestos, penalización de carreras laborales discontinuas o estereotipos ocupacionales–, el sistema tiende a aprender y reproducir esas desigualdades, extendiéndolas a nuevas versiones y ex-

⁵⁹ [https://www.artificial-intelligence-act.com/Artificial_Intelligence_Act_Article_10_\(Proposal_25.11.2022\).html](https://www.artificial-intelligence-act.com/Artificial_Intelligence_Act_Article_10_(Proposal_25.11.2022).html)

portándolas a contextos distintos. De ahí que la transparencia y la trazabilidad de los datos sean condiciones indispensables para detectar y corregir sesgos: si no se puede reconstruir qué información alimentó el modelo ni cómo se transformó, resulta ilusorio pretender una corrección técnica eficaz.

Por esta razón, más allá de analizar exclusivamente las métricas técnicas de rendimiento del modelo, es imprescindible interrogar los criterios que se automatizan. Los criterios abarcan tanto lo que se decide automatizar (por ejemplo, la “idoneidad” de una persona candidata), como los objetivos de optimización (minimizar costes, maximizar productividad, reducir rotación) y las variables consideradas relevantes (historial de empleo, horarios, bajas médicas, etc.), que en conjunto determinan qué resultados privilegia el sistema. No es posible ajustar parámetros ni “corregir” un modelo sin cuestionar previamente estas decisiones de diseño, pues son ellas las que pueden incrustar y legitimar sesgos de género y otros sesgos estructurales. La literatura sobre automatización como intervención “antiseggo” ha mostrado, además, que las soluciones puramente técnicas –por ejemplo, reequilibrar pesos o aplicar filtros de *fairness*– son insuficientes si no se revisan los supuestos normativos que guían la construcción de los criterios automatizados⁶⁰.

Comprender la naturaleza real del sesgo algorítmico –incluido su componente institucional y estructural– no sólo es condición para mitigarlo en un sistema concreto, sino también presupuesto para diseñar marcos jurídicos eficaces. Para regular de forma adecuada la automatización en ámbitos sensibles como el empleo y garantizar el derecho a la igualdad, resulta necesario ir más allá del ajuste técnico de modelos y abordar cuestiones como quién define los criterios de evaluación, qué valores organizativos se priorizan y cómo se distribuyen las responsabilidades entre proveedores, empleadores y otros agentes involucrados. Sin esta perspectiva,

⁶⁰ <https://artificialintelligenceact.eu/annex/4/>

el Derecho corre el riesgo de limitarse a exigir métricas de transparencia formales que no desvelan las decisiones valorativas que estructuran el pipeline.

La noción de “*pipeline*” de aprendizaje automático ayuda a entender la profundidad de este problema. Un pipeline se define como una cadena estructurada de pasos que transforma datos brutos en predicciones operativas mediante diferentes capas funcionales. Entre sus componentes típicos se encuentran: una capa de datos, encargada de gestionar el acceso a las fuentes de información y garantizar su trazabilidad temporal; una capa de “*features*”, que convierte datos crudos en variables de entrada para el modelo; una capa de “*scoring*”, que toma el vector de características y produce las predicciones o decisiones automatizadas; y una capa de evaluación, que mide, compara y valida el comportamiento de los modelos para asegurar su rendimiento y detectar desviaciones. En teoría, esta estructura modular debería facilitar la reproducibilidad y la estandarización de las decisiones técnicas, permitiendo auditar qué se decidió en cada fase⁶¹.

En la práctica, sin embargo, muchos sistemas complejos no se ajustan a este ideal. La ausencia de especificaciones claras y de límites robustos entre componentes conduce a pipelines heterogéneos, con capas difusas e inestables. El comportamiento del sistema pasa a depender en gran medida de la distribución de los datos, que evoluciona con el tiempo, y la preparación de datos suele construirse de forma incremental, acumulando pasos fragmentarios y grandes cantidades de “código pegamento” que conecta librerías y paquetes sin una arquitectura claramente documentada. Esta acumulación de decisiones técnicas dispersas genera lo que se ha denominado una “selva de pipelines”, en la que desaparecen los límites nítidos entre capas, se dificulta la actualización segura de los modelos y se complica enormemente la trazabilidad de las decisiones.

⁶¹ <https://www.glocertinternational.com/resources/guides/eu-ai-act-technical-documentation-article-11/>

Cuando falta trazabilidad, no existe un registro sistemático del origen de los datos, de las etiquetas utilizadas para entrenar el modelo, de las transformaciones aplicadas ni de las decisiones de diseño adoptadas en cada fase del pipeline. Todas estas decisiones –qué se etiqueta como “éxito”, qué variables se crean o se descartan, cómo se imputan valores ausentes, qué función objetivo se optimiza– son, en realidad, decisiones normativas que incorporan criterios valorativos sobre qué conductas se premian o penalizan. En tales condiciones, el pipeline se convierte en una auténtica “caja negra” sociotécnica que oculta tanto los juicios de valor como las opciones técnicas subyacentes, impidiendo una auditoría legal efectiva, dificultando la atribución de responsabilidades y bloqueando cualquier corrección estructural de los sesgos presentes en el sistema.

Precisamente esta situación es la que el Reglamento de Inteligencia Artificial pretende evitar al exigir documentación detallada de los datos, de los modelos y de las decisiones tomadas a lo largo de todo el ciclo de vida del sistema de alto riesgo. La norma aspira a que queden recogidas de forma explícita las elecciones técnicas y valorativas que se adoptan en el pipeline –desde la selección de datos hasta la definición de métricas de rendimiento y equidad–, porque son esas decisiones las que determinan los riesgos jurídicos y de discriminación y sólo pueden ser controladas si son rastreables y verificables por autoridades, personas afectadas y representantes de los trabajadores. Sin trazabilidad y sin transparencia sustantiva del pipeline, cualquier promesa de “IA responsable” corre el riesgo de convertirse en mera retórica, dejando sin protección efectiva el derecho a la igualdad y a la no discriminación en contextos tan sensibles como el empleo y la protección social⁶².

⁶² <https://www.glocertinternational.com/resources/guides/eu-ai-act-technical-documentation-article-11/>

4.4. GOBERNANZA DE DATOS EN LA IA DEL ALTO RIESGO: DE LA "IA CENTRADA EN EL MODELO" A LA "IA CENTRADA EN LOS DATOS" PARA PREVENIR SESGOS Y DISCRIMINACIÓN

Ante las necesidades señaladas, el Reglamento europeo de inteligencia artificial introduce para los sistemas de alto riesgo un conjunto de obligaciones que trasciende el mero diseño del modelo y se proyecta sobre todo el ciclo de vida de los datos: documentación técnica exhaustiva, registro de eventos, vigilancia posterior a la comercialización y, de forma muy destacada, requisitos reforzados de calidad de los datos, documentación del etiquetado y registros automáticos que permitan reconstruir el proceso de decisión.

En primer lugar, y antes de analizar específicamente los riesgos de discriminación, resulta necesario detenerse en los riesgos asociados a la calidad de los datos, dada su estrecha conexión. El Reglamento se centra en regular la entrada de datos en los sistemas de inteligencia artificial de alto riesgo, partiendo de la premisa de que muchos de los problemas de seguridad, fiabilidad y equidad se incuban en esta fase. Ya antes de su aprobación, parte de la literatura jurídica había advertido de que la atención normativa se estaba concentrando en los resultados y aplicaciones de la IA⁶³., descuidando los datos de entrenamiento, pese a que numerosas fuentes de riesgo –entre ellas los sesgos de género– tienen precisamente su origen en esos datos.

En el ámbito técnico se formulaba una crítica paralela: la investigación en detección de intrusiones y en otros usos de la IA se había centrado en diseñar y optimizar modelos, relegando a un segundo plano el estudio sistemático de los datos, sus requisitos de calidad y las garantías necesarias para que fueran adecuados al entrenar y evaluar sistemas seguros y fiables. Otros autores fueron más específicos al diagnosticar que la IA se había con-

⁶³ TRAN, N. et al., "Data Curation and Quality Evaluation for Machine Learning-Based Cyber Intrusion Detection", IEEE Access 10 (2022), pp. 1-2, en <https://ieeexplore.ieee.org/ielam/6287639/9668973/9907008-aam.pdf>

figurado como un campo “centrado en el modelo”, en el que el objetivo principal era mejorar arquitecturas y algoritmos mientras se mantenían prácticamente invariables los conjuntos de datos⁶⁴.

Este paradigma llegó a su límite cuando se comprobó empíricamente que perfeccionar el modelo ya no bastaba para obtener mejoras sustanciales en problemas del mundo real. Al hacerse evidente que la provisión y la selección de datos podían afectar positiva o negativamente a la eficacia y eficiencia del modelo, comenzó a consolidarse un giro hacia una “IA centrada en los datos”, que investiga de manera prioritaria cómo la cantidad, la calidad y la representatividad de los datos condicionan el comportamiento de los sistemas. En este enfoque, la mejora de los datos se entiende como un factor decisivo para el éxito de los sistemas de IA, en igualdad de importancia –o incluso por encima– de la sofisticación del modelo.

Subestimar la calidad de los datos y su impacto en el comportamiento del modelo en contextos reales puede desencadenar lo que se ha denominado “cascadas de datos”: procesos acumulativos en los que errores, carencias o sesgos presentes en las primeras fases del ciclo de datos se transmiten y amplifican aguas abajo, generando deuda técnica y degradando progresivamente el sistema. Estas cascadas producen, al menos, dos tipos de efectos negativos. Por un lado, una disminución de la precisión y del rendimiento de los modelos, al estar entrenados con datos incompletos, ruidosos o mal etiquetados; por otro, la cristalización de sesgos persistentes que se traducen en decisiones discriminatorias reproducibles, difíciles de detectar y corregir una vez que el sistema está desplegado.

Desde la óptica de la igualdad y la no discriminación, una de las utilidades más relevantes de la IA centrada en los datos es preci-

⁶⁴ ZHA, D. et al., “Data-Centric Artificial Intelligence: A Survey”, (2023), p. 4. <https://arxiv.org/pdf/2303.10158>

samente la identificación y corrección de fuentes de discriminación en los conjuntos de datos. Numerosos estudios han puesto de manifiesto que sistemas utilizados en ámbitos de alto riesgo –como la selección de personal, la concesión de créditos o la evaluación de prestaciones sociales– exhiben comportamientos discriminatorios frente a determinados grupos, en particular mujeres, personas migrantes o minorías étnicas. Esta falta de equidad suele estar asociada a sesgos de representación: el dataset sobre el que aprende el modelo contiene una infrarrepresentación o una representación distorsionada de ciertos colectivos, de modo que el sistema aprende a partir de muestras parciales y acaba ofreciendo un rendimiento sistemáticamente peor –o decisiones más desfavorables– para esos grupos.

La perspectiva centrada en los datos sostiene que la forma más sólida de afrontar este problema es garantizar que los datos sean representativos, limpios y útiles, y que su gestión incorpore mecanismos específicos de mitigación de sesgos. En términos operativos, ello implica, al menos, tres líneas de actuación. En primer lugar, desarrollar procedimientos de depuración y enriquecimiento de datos que corrijan desequilibrios de representación y errores de etiquetado, especialmente allí donde las trayectorias laborales de las mujeres o de otros colectivos vulnerables aparecen fragmentadas o infravaloradas. En segundo lugar, diseñar metodologías y protocolos de evaluación que permitan revelar, mediante análisis estadístico y pruebas de equidad, si un sistema de IA⁶⁵ genera trato desigual entre grupos comparables, identificando disparidades en tasas de selección, en recomendaciones o en resultados. Por último, establecer procesos de monitorización continua que detecten y corrijan de forma iterativa las desigualdades en los datos a lo largo del tiempo, teniendo en cuenta el carácter dinámico de los contextos en los que los sistemas se utilizan (cambios en

⁶⁵ GUDIVADA, V., APON, A. y DING, J., “Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations”, *International Journal on Advances in Software*, vol. 10, no. 1&2, pp. 3-5, 2017.

el mercado de trabajo, nuevas formas de empleo, modificaciones normativas, etc.).

En definitiva, el giro desde una IA centrada en el modelo hacia una IA centrada en los datos, que el Reglamento europeo de inteligencia artificial recoge y refuerza mediante exigencias de calidad, documentación y trazabilidad, no es solo una cuestión técnica. Tiene una dimensión claramente jurídica y de derechos fundamentales: sin datos de calidad, representativos y auditables, la prevención de sesgos y de discriminación –incluidos los de género– se convierte en un objetivo meramente retórico. Por el contrario, una gobernanza robusta de los datos de entrenamiento, validación y prueba constituye la condición de posibilidad para que las obligaciones de no discriminación se traduzcan en sistemas de IA de alto riesgo que no reproduzcan, ni amplifiquen, las desigualdades ya existentes.

4.5 ENTRE "FAIRNESS" Y NO DISCRIMINACIÓN: LÍMITES DEL REGLAMENTO DE IA PARA PROTEGER EL DERECHO A LA IGUALDAD

Tratar de encajar el Reglamento de Inteligencia Artificial en el marco del derecho antidiscriminatorio resulta complejo porque las categorías centrales de uno y otro campo no son plenamente equivalentes. En el ámbito de la informática, la equidad suele entenderse como un conjunto de requisitos técnicos destinados a evitar resultados moralmente indeseables de los algoritmos, mientras que en el lenguaje jurídico el término relevante no es la equidad, sino el derecho fundamental a la igualdad y a la no discriminación. De este desajuste terminológico se sigue que conceptos clave para regular el sesgo algorítmico –como *fairness*, *bias mitigation* o individual *fairness*– no se corresponden sin más con las categorías dogmáticas que utiliza el derecho, lo que dificulta construir puentes directos entre métricas técnicas y estándares jurídicos de igualdad.

Tanto la ingeniería como el derecho persiguen, en última instancia, evitar que los sistemas de IA produzcan resultados discriminatorios, pero lo hacen desde enfoques distintos. En la literatura técnica, la detección, prevención y mitigación de desequilibrios se articula mediante métricas, fórmulas matemáticas y reglas estadísticas que operan, sobre todo, sobre la entrada de los sistemas (datos, etiquetas, distribuciones de probabilidad). El derecho antidiscriminatorio, en cambio, se centra en el resultado de una práctica concreta, valorando si la decisión impugnada vulnera la normativa de igualdad en atención al caso y al contexto: el análisis se realiza *ex post*, sobre la base de efectos observables y contrastables ante los tribunales. De ahí que la no discriminación y la igualdad no puedan reducirse a un problema estadístico ni a la mera ausencia de sesgo en los datos, sino que constituyen conceptos normativos que implican juicios de legitimidad sobre el trato diferencial.

La jurisprudencia constitucional española⁶⁶ ha subrayado que, aunque el principio general de igualdad admite diferencias de trato justificadas por criterios de razonabilidad y proporcionalidad⁶⁷, el derecho fundamental a la no discriminación exige un control mucho más intenso. En este ámbito, no basta con que la diferencia de trato sea razonable: se aplica un canon reforzado que permite presumir la irrazonabilidad de la diferenciación cuando afecta a motivos especialmente protegidos (entre ellos, el sexo o el género), de modo que sólo se admite el trato desigual como excepción y bajo condiciones estrictas de proporcionalidad. Esta lógica se distancia de muchas propuestas de *fairness* algorítmica, que tienden a considerar aceptables determinadas disparidades siempre que se mantengan dentro de umbrales técnicos, sin incorporar necesariamente una valoración jurídica del impacto sobre grupos históricamente discriminados.

⁶⁶ SSTC 233/2007 de 5 de noviembre, FJ 5.

⁶⁷ MONEREO PÉREZ, J. L., “Vulneración del derecho a la no discriminación y principio de interpretación jurídica conforme a la eficacia de la tutela antidiscriminatoria: salarios de tramitación y despido nulo por embarazo”, *Revista de Jurisprudencia Laboral* 4 (2024).

Por ello, identificar sin más sesgo con discriminación sería equivocarse⁶⁸. El sesgo es, en términos amplios, una desviación sistemática en los datos o en el funcionamiento del modelo; la discriminación, en cambio, es una calificación jurídica que se proyecta sobre el resultado y su compatibilidad con los derechos fundamentales. El Reglamento de Inteligencia Artificial asume, en esencia, que un proveedor puede adoptar medidas técnicas para mitigar sesgos (mejorando calidad de datos, aplicando técnicas de *debiasing*, ajustando modelos)⁶⁹, pero ello no garantiza que el sistema no produzca resultados discriminatorios en el sentido jurídico del término. Cumplir formalmente con los requisitos de calidad de los datos y con distintas técnicas de corrección estadística no equivale, por sí solo, a respetar el derecho a la igualdad y a la no discriminación.

En este contexto se han desarrollado múltiples métricas de equidad que pretenden adecuarse a las particularidades de cada sistema –por ejemplo, igualdad de oportunidades, paridad demográfica o igualdad de falsos positivos–. Sin embargo, estas métricas son a menudo incompatibles entre sí y generan resultados contradictorios; además, se ha mostrado que, en muchos casos, no satisfacen las exigencias del derecho antidiscriminatorio, que no se conforma con un equilibrio estadístico abstracto, sino que exige examinar quién soporta los impactos negativos y si estos refuerzan desigualdades estructurales, como las de género. En consecuencia, puede haber sistemas técnicamente “justos” según una métrica de fairness y, sin embargo, jurídicamente discriminatorios, porque producen desventajas sistemáticas para mujeres u otros colectivos protegidos.

Desde esta perspectiva, puede afirmarse que los sesgos y las lesiones de derechos son categorías relacionadas pero diferencia-

⁶⁸ MAYORALAS TERRAZAS I., “Inteligencia artificial y sus sesgos en el análisis de datos: los riesgos de la nueva discriminación” en *Ensayos Ciberfeministas* (Dykinson, 2023), pp. 227-229.

⁶⁹ El considerando 45 del Reglamento de Inteligencia Artificial establece que el mismo no debe interferir a la aplicación del derecho antidiscriminatorio ya existente en la Unión Europea.

das. Los sesgos constituyen la causa –desviaciones en los datos, en los modelos o en los procesos de decisión–, mientras que la discriminación y los daños derivados son el efecto cuando dichas desviaciones inciden sobre derechos protegidos, como el derecho a la igualdad y a la no discriminación por razón de género. El Reglamento de IA se orienta principalmente a regular medidas técnicas para disminuir el sesgo, mientras que la normativa antidiscriminatoria, que permanece plenamente aplicable según el considerando 45 del propio Reglamento, se ocupa de valorar y sancionar los resultados discriminatorios.

De este modo, el Reglamento de Inteligencia Artificial proporciona un conjunto de obligaciones y herramientas –dirigidas a proveedores, usuarios y demás agentes– que pueden servir para identificar y reducir el riesgo de que los sistemas adopten conductas sesgadas que, desde la óptica del derecho, terminarían siendo discriminatorias. Su enfoque es eminentemente preventivo: se centra en la identificación temprana de sesgos y en su mitigación a nivel técnico, con la finalidad de disminuir la probabilidad de resultados vulneradores del derecho a la igualdad. Ahora bien, junto a estas medidas preventivas, resulta esencial reforzar las exigencias de transparencia y explicabilidad, especialmente una vez desplegados los sistemas, para afrontar el problema de la opacidad de los algoritmos complejos. Solo con transparencia suficiente –sobre datos, modelos y criterios de decisión– será posible vincular concretamente los casos de discriminación detectados en la práctica con el funcionamiento de sistemas de IA y, en su caso, activar eficazmente los mecanismos de responsabilidad previstos por el derecho antidiscriminatorio.

4.6 DERECHO A LA EXPLICACIÓN COMO LLAVE PARA DENUNCIAR LA DISCRIMINACIÓN ALGORÍTMICA

La efectividad del derecho a la no discriminación se enfrenta a obstáculos significativos, especialmente para las personas traba-

jadoras que intentan impugnar un trato presuntamente discriminatorio vinculado al uso de sistemas de inteligencia artificial. La principal dificultad reside en probar ante el órgano judicial cuál ha sido la causa real de la decisión, lo que en el contexto de modelos opacos y de datos inaccesibles se convierte, en la práctica, en una tarea casi imposible. Cuando la posible discriminación – incluida la de género– está mediada por un sistema algorítmico, la opacidad técnica y la falta de acceso a los datos agravan las asimetrías probatorias y reducen drásticamente la capacidad de las víctimas para identificar y demostrar el sesgo subyacente.

Para paliar en parte esta situación, el artículo 86 del Reglamento de Inteligencia Artificial reconoce un derecho a obtener explicaciones sobre decisiones individuales adoptadas con base en sistemas de IA de alto riesgo. La comprensibilidad y la transparencia de los criterios utilizados para decidir, así como de las conclusiones extraídas, son esenciales para que la persona afectada pueda saber cómo se ha aplicado el sistema a su caso concreto y qué elementos debe trasladar al órgano judicial para sustentar su demanda. Sin una explicación mínimamente inteligible, la víctima desconoce tanto el papel que ha desempeñado el algoritmo como los factores que han influido en el resultado, lo que vacía de contenido práctico el derecho a la tutela antidiscriminatoria.

El artículo 86.1 dispone que toda persona afectada por una decisión adoptada por el responsable del despliegue basándose en la salida de un sistema de IA de alto riesgo incluido en el anexo III –salvo las excepciones de su punto 2– y que produzca efectos jurídicos o le afecte de manera significativa, con impacto potencial sobre su salud, seguridad o derechos fundamentales, tiene derecho a obtener explicaciones claras y significativas sobre el papel desempeñado por el sistema y sobre los elementos principales de la decisión. Este derecho se aplica, entre otros, a los sistemas utilizados en el ámbito del empleo, la gestión de personas trabajadoras y el acceso al autoempleo, así como a los sistemas

empleados en la garantía del cumplimiento del Derecho, siempre que su uso esté permitido por la normativa de la Unión o nacional. En comparación con el artículo 13.2.f) del RGPD, que ofrecía una base más genérica para reclamar información sobre la lógica del tratamiento automatizado, el artículo 86.1 supone un avance específico y reforzado en términos de transparencia para los sistemas de alto riesgo.

La obligación derivada de este precepto consiste en proporcionar información “clara y significativa” sobre la lógica utilizada por el sistema, lo que resulta especialmente exigente cuando se trata de algoritmos complejos o de “caja negra”. Aunque el Reglamento presupone que quienes despliegan el sistema deben ofrecer el máximo de información posible sobre el proceso de decisión, la realidad técnica impone límites: cuanto mayor es la autonomía y complejidad del modelo, mayor dificultad existe para describir de manera comprensible su funcionamiento interno y su contribución específica al resultado.

Con todo, el derecho a la explicación se erige en un requisito previo indispensable para poder impugnar jurídicamente una decisión algorítmica. Solo si la persona afectada entiende, siquiera de forma aproximada, en qué se ha basado la decisión, podrá cuestionar ante los tribunales si ese resultado vulnera su derecho a la igualdad y a la no discriminación, por ejemplo, por haber ponderado de manera desfavorable variables estrechamente asociadas al género. Sin embargo, la efectividad de este derecho sigue dependiendo de un nivel mínimo de explicabilidad: el artículo 86.1 solo puede cumplirse si el sistema está diseñado y documentado de forma que permita generar explicaciones comprensibles, lo que conecta de nuevo con las obligaciones de gobernanza de datos, transparencia y documentación técnica que el propio Reglamento impone a lo largo de todo el ciclo de vida de la inteligencia artificial.

4.7 SANCIÓN Y REPARACIÓN FRENTE A LA DISCRIMINACIÓN ALGORÍTMICA: ENTRE LA TUTELA Y LAS NUEVAS FACULTADES DEL REGLAMENTO DE IA

Según el artículo 17 de la Directiva 2000/78/CE sobre igualdad en el empleo, el incumplimiento de la normativa antidiscriminatoria debe ir acompañado de sanciones efectivas, proporcionadas y disuasorias. En la práctica, los sistemas nacionales de ejecución en la Unión Europea se han configurado, en su mayoría, con un marcado carácter individualista y reparador, más orientado a resarcir a la víctima que a prevenir estructuralmente futuras vulneraciones. En el ordenamiento español, el Tribunal Supremo ha subrayado que, en numerosos casos⁷⁰, la cuantificación prevista en la Ley sobre Infracciones y Sanciones en el Orden Social (LISOS) no resulta suficiente para cumplir simultáneamente las funciones de reparación del daño y de disuasión de nuevas lesiones de derechos fundamentales, debido a la amplitud excesiva de las horquillas sancionadoras. Por ello, aun tomando como referencia las sanciones tipificadas en la LISOS, los tribunales deben valorar siempre las circunstancias concretas del caso, incluida la intensidad del quebrantamiento del derecho y las consecuencias que la vulneración ha tenido en la situación personal y social de la persona afectada.

En materia indemnizatoria, el artículo 183.1 y 2 de la Ley reguladora de la jurisdicción social (LRJS) establece que, cuando exista discriminación u otra vulneración de derechos fundamentales, el órgano judicial debe fijar una indemnización que abarque tanto el daño moral como otros daños y perjuicios derivados de la lesión⁷¹. La norma añade que la determinación de la cuantía ha de hacerse con prudencia, en especial cuando resulte muy difícil resarcir íntegramente a la víctima. A partir de este marco, el Tribunal Supremo⁷² ha configurado el resarcimiento del daño moral

⁷⁰ STS de 24 junio 2024 (rec. 1113/2024); STS de 13 de julio de 2015 (rec. 221/2014).

⁷¹ STSJ de la Comunidad Valenciana (de 8 de octubre de 2025 (rec. 1389/2025).

⁷² Véase STS 11 de junio de 2012 (rec. 3336/11).

con un criterio abierto: considera imprescindible reconocer un margen de discrecionalidad judicial para traducir el sufrimiento de la víctima en términos económicos, al no existir parámetros objetivos que permitan hacerlo de manera automática. Cuando el daño moral está directamente vinculado a la vulneración de un derecho fundamental, los jueces pueden realizar una valoración más global y menos formalista, adoptando valoraciones flexibles y contextualizadas que reflejen mejor la realidad del caso.

La Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, refuerza esta orientación reparadora. Su artículo 27 impone a la persona física o jurídica que incurra en discriminación la obligación de reparar el daño mediante indemnización y, cuando sea posible, mediante la restitución de la situación anterior para la víctima. Además, dispone que, una vez acreditada la discriminación, la existencia de daño moral ha de presumirse, debiendo valorarse en función de las circunstancias del caso. La jurisprudencia del Tribunal Superior de Justicia de la Comunidad Valenciana ha precisado que, en tales supuestos, la reparación debe articularse en dos planos: por un lado, la restitución de la situación previa a la conducta discriminatoria; por otro, la indemnización económica. Asimismo, se entiende que el daño moral se presume en toda víctima de discriminación y que su cuantificación debe tener en cuenta la posible concurrencia de varias causas de discriminación y la gravedad de la lesión, otorgando especial relevancia al medio a través del cual se ha producido la conducta discriminatoria, aspecto clave cuando el instrumento de decisión ha sido un sistema de inteligencia artificial.

En cuanto a la determinación de la cuantía solicitada, el artículo 179.3 LRJS impone a la parte demandante la obligación de exponer en la demanda las circunstancias relevantes para fijar la indemnización –gravedad, duración, consecuencias del daño, bases de cálculo de los perjuicios–, salvo en los casos de daños morales asociados a la vulneración de derechos fundamentales cuando su estimación resulte especialmente difícil. El Tribunal Supremo ha

señalado que los daños morales se hallan íntimamente conectados con la lesión de dichos derechos y que, dada la complejidad de su cuantificación, no puede exigirse a la persona demandante que aporte bases exactas para concretar la cuantía. Esto se traduce en un mayor margen de apreciación judicial a la hora de valorar el daño moral, margen particularmente relevante en contextos de discriminación algorítmica en los que el impacto lesivo se manifiesta de forma difusa pero profunda.

Sobre este telón de fondo, la doctrina constitucional distingue entre derechos constitucionales específicamente laborales –por ejemplo, el derecho a la negociación colectiva o el derecho al trabajo– y derechos constitucionales de carácter general que despliegan una dimensión específica en el ámbito laboral cuando se ejercen en el marco de la relación de trabajo. Estos últimos, denominados derechos laborales inespecíficos⁷³, incluyen el derecho a la igualdad y a la no discriminación, cuya protección en el ámbito laboral se apoya en normas de carácter general, en gran medida de inspiración civil, que no siempre ofrecen un nivel de especialización adecuado a la complejidad de las desigualdades en el empleo. Como consecuencia, el ordenamiento laboral español dispone de un cuerpo normativo relativamente escaso en materia de igualdad, lo que incrementa la dependencia de la jurisprudencia –tanto interna como europea– para concretar el alcance de estos derechos, circunstancia especialmente delicada cuando entran en juego procesos automatizados de decisión.

A ello se suma la primacía del Derecho de la Unión Europea⁷⁴. El artículo 93 de la Constitución española y el artículo 2 del Tratado de Funcionamiento de la Unión Europea reconocen la prevalencia del Derecho de la Unión en las materias de su competencia, obligando a los tribunales nacionales a interpretar y aplicar el de-

⁷³ PALOMEQUE LÓPEZ, Los derechos laborales en la Constitución Española, *Cuadernos y Debates* 28, p. 31.

⁷⁴ Declaración del Pleno del Tribunal Constitucional 1/2004, del 13 de diciembre de 2004.

recho interno de conformidad con la jurisprudencia del Tribunal de Justicia de la Unión Europea (TJUE). Aunque el TJUE no es formalmente un tribunal de derechos humanos, los derechos fundamentales han ocupado un lugar central en su jurisprudencia, especialmente a partir de las directivas sobre igualdad de trato. Esta influencia refuerza la exigencia de que las sanciones e indemnizaciones frente a la discriminación –incluida la de carácter algorítmico– sean efectivas y disuasorias, en coherencia con el artículo 17 de la Directiva 2000/78/CE.

El Reglamento de Inteligencia Artificial se inserta en este entramado sin alterar los mecanismos nacionales de reparación del derecho a la no discriminación: en los casos de discriminación algorítmica, deben seguirse las vías procesales ya existentes. Sin embargo, el Reglamento introduce mecanismos complementarios orientados a reforzar la tutela efectiva. En particular, su artículo 77 se dirige a las autoridades y organismos públicos nacionales competentes en materia de derechos fundamentales –incluido el derecho a la no discriminación– para precisar sus facultades en el ámbito de los sistemas de IA de alto riesgo. El artículo 77.1 les reconoce la facultad de solicitar cualquier documentación creada o conservada en aplicación del Reglamento de IA y de acceder a ella “en un lenguaje y formato accesibles”, siempre que dicho acceso sea necesario para el ejercicio efectivo de sus mandatos. El apartado 2 obliga a los Estados miembros a identificar y comunicar estas autoridades a la Comisión y a los demás Estados, mientras que el apartado 3 permite que dichas autoridades soliciten a la autoridad nacional de vigilancia del mercado la realización de pruebas sobre sistemas de IA de alto riesgo cuando la documentación disponible no sea suficientemente reveladora de una posible vulneración de derechos fundamentales.

No obstante, estas facultades presentan limitaciones significativas. En primer lugar, su ámbito de aplicación se restringe a los sistemas de IA de alto riesgo del anexo III, de modo que quedan fuera otros sistemas que, aun sin alcanzar ese umbral, pueden generar efectos discriminatorios relevantes. En segundo lugar, el Reglamento no

especifica qué tipo de pruebas pueden solicitar las autoridades, lo que genera incertidumbre interpretativa; si bien el capítulo VI contempla distintos regímenes de pruebas en entornos controlados y en condiciones reales, la falta de remisión expresa sugiere que el abanico es amplio, pero también poco definido.

Además, existe una asimetría entre las obligaciones de los proveedores frente a las autoridades de vigilancia del mercado –reguladas en el artículo 21 RIA– y la posición de otras autoridades de derechos fundamentales. Mientras los proveedores están obligados a facilitar a las primeras toda la información necesaria para acreditar la conformidad del sistema, no existe un deber análogo de suministro directo a las segundas, que deben canalizar sus peticiones a través del mecanismo del artículo 77. Dado que las autoridades de vigilancia del mercado difieren entre Estados miembros en cuanto a recursos y capacidades, pueden generarse niveles de protección desiguales; la fuerte dependencia de la autoevaluación de los proveedores –materializada en informes y declaraciones de conformidad– limita, además, la capacidad de control externo efectivo.

Las dificultades se intensifican si se tiene en cuenta que las autoridades nacionales carecen, en la actualidad, de estándares objetivos detallados para evaluar la conformidad de los sistemas de IA con el Reglamento, lo que reduce la previsibilidad y coherencia de sus actuaciones. Aunque las autoridades de protección de datos cuentan con experiencia relevante, existe el riesgo de sobrecarga y falta de recursos si concentran la mayoría de las reclamaciones relacionadas con IA y discriminación. En el ámbito procesal, la persona demandante que alegue discriminación algorítmica debe afrontar una carga probatoria inicial particularmente exigente: la demostración del nexo causal entre el trato desfavorable y la característica protegida requiere habitualmente de pruebas complejas, como análisis estadísticos y periciales sobre el funcionamiento del sistema. Aun cuando el Reglamento reconoce derechos de acceso a información y documentación, la protección de secretos comerciales y de pro-

piedad intelectual limita el nivel de detalle que puede obtenerse, lo que puede convertir esos derechos en herramientas insuficientes.

El acceso íntegro al código fuente o al algoritmo suele ser incompatible con la tutela del secreto empresarial. El considerando 63 del RGPD fija un estándar mínimo de equilibrio entre el derecho de acceso a la información y la protección de secretos comerciales, señalando que esta última no puede invocarse como excusa para negar toda información sobre la lógica aplicada en decisiones automatizadas. Por su parte, el artículo 78 RIA protege la confidencialidad de la información obtenida en aplicación del Reglamento –incluido el código fuente–, pero remite a las excepciones del artículo 5 de la Directiva (UE) 2016/943, que permite limitar la protección del secreto empresarial cuando la divulgación sea necesaria para el ejercicio legítimo de determinadas funciones o derechos.

La Directiva 2016/943 ha sido transpuesta al Derecho español mediante la Ley 1/2019, de secretos empresariales, cuyo artículo 2 considera lícita la obtención, utilización o revelación del secreto cuando se realice, entre otros supuestos, en el ejercicio de los derechos de información y consulta de las personas trabajadoras y sus representantes. Esta previsión se complementa con el artículo 1.3 de la misma Ley, que impide que la protección del secreto empresarial menoscabe la autonomía de los interlocutores sociales o el derecho a la negociación colectiva. En conjunto, estas normas abren una vía para que la negociación colectiva y los derechos de información, consulta y participación funcionen como contrapeso a la opacidad de los sistemas de IA utilizados en la gestión de personal, convirtiéndose en herramientas clave para la defensa de los derechos fundamentales –en particular, el derecho a la igualdad y a la no discriminación de las mujeres– frente a los riesgos de la discriminación algorítmica.

5. EXPERIENCIAS COMPARADAS EN LA REGULACIÓN DE LA IA LABORAL CON PERSPECTIVA DE IGUALDAD

La progresiva incorporación de sistemas de inteligencia artificial y herramientas algorítmicas a los procesos de gestión de recursos humanos no es un fenómeno exclusivamente europeo, sino que ha motivado respuestas regulatorias incipientes en diversos ordenamientos, particularmente en el ámbito anglosajón. Estas respuestas, aun fragmentarias, ofrecen indicios relevantes sobre las posibles vías de intervención jurídica frente a la discriminación algorítmica en el empleo y permiten contrastar el modelo europeo, fuertemente apoyado en el Derecho de la Unión Europea y en la dogmática antidiscriminatoria, con aproximaciones más sectoriales o basadas en el *enforcement* administrativo. Dentro de este panorama comparado, resultan especialmente significativas, de un lado, algunas normas subnacionales de Estados Unidos que abordan de forma específica el uso de algoritmos en la selección de personal y, de otro, las primeras experiencias de autoridades con competencias laborales o antidiscriminatorias que han comenzado a intervenir sobre la IA con un enfoque explícito de igualdad.

5.1. REGULACIONES SUBNACIONALES EN ESTADOS UNIDOS SOBRE ALGORITMOS DE SELECCIÓN Y AUDITORÍAS DE SESGO

En el ordenamiento estadounidense, la ausencia de una legislación federal integral sobre inteligencia artificial se ha visto parcialmente compensada por iniciativas estatales y locales orientadas a controlar el uso de sistemas algorítmicos en la selección de personal. Estas normas se insertan sobre el sustrato de la legisla-

ción antidiscriminatoria federal (particularmente el Title VII de la Civil Rights Act) y de la doctrina del *disparate impact*, pero introducen obligaciones específicas de transparencia, auditoría y consentimiento cuando se emplean herramientas automatizadas en decisiones de empleo.

Un ejemplo paradigmático lo constituye la Local Law 144/2021 de la ciudad de Nueva York⁷⁵, relativa a los denominados *Automated Employment Decision Tools* (AEDT). Esta norma prohíbe a empleadores y agencias de empleo utilizar herramientas automatizadas de decisión en materia laboral –incluyendo, entre otros, sistemas de cribado de candidaturas, clasificación de perfiles o recomendaciones de contratación y promoción– salvo que se cumplan tres requisitos acumulativos: a) realización de una auditoría de sesgo por un tercero independiente en el año anterior a su utilización; b) publicación en la página web de la empresa de un resumen de los resultados de la auditoría; y c) suministro de información previa a las personas candidatas y trabajadoras sobre el uso de la herramienta y los datos que serán tratados.

El concepto de *bias audit* se define como una evaluación independiente destinada, al menos, a medir el posible impacto dispar de la herramienta sobre distintos grupos protegidos, incluyendo el sexo, la raza y la etnia, mediante el cálculo de ratios de impacto y otras métricas similares. La auditoría debe analizar si el sistema produce tasas significativamente diferentes de selección o recomendación entre hombres y mujeres, así como entre otros grupos comparables, y exige que se justifique la exclusión de determinadas categorías cuando su tamaño muestral resulte insuficiente para un análisis estadístico robusto. La lógica subyacente se aproxima, por tanto, a la de la discriminación indirecta: se presumen problemáticos aquellos resultados que generan una desventaja significativa para un grupo protegido, salvo que el empleador pueda acreditar que el instrumento es necesario y está

⁷⁵ https://www.nycbiasaudit.com/?web_page_name=%2F

estrechamente vinculado a las exigencias del puesto, en términos próximos al estándar de *job related and consistent with business necessity* desarrollado por la jurisprudencia estadounidense.

La Local Law 144 introduce además obligaciones de transparencia individual y colectiva. Desde la perspectiva individual, el empleador debe informar a la persona candidata, con antelación razonable, de que se va a utilizar un AEDT en el proceso de selección, del tipo de datos que se van a recabar y de los criterios generales de funcionamiento del sistema, lo que guarda evidente paralelismo con las obligaciones de información del RGPD y con el art. 22 respecto de decisiones automatizadas. En el plano colectivo, la exigencia de publicar un resumen de la auditoría refuerza no solo la rendición de cuentas frente a la opinión pública, sino también la posibilidad de que sindicatos, organizaciones de derechos civiles y oficinas de igualdad analicen de forma sistemática los patrones de impacto dispar de estas herramientas, facilitando la detección de discrimen estructural.

Aunque la norma neoyorquina no se presenta como un instrumento específicamente laboral, su ámbito material se centra en decisiones de empleo y promoción, y el contenido de las obligaciones de auditoría se orienta explícitamente a controlar la discriminación algorítmica en el acceso al trabajo, con mención expresa a las categorías de sexo, raza y etnia como ejes de análisis. En este sentido, la Local Law 144 anticipa, a nivel local, exigencias que recuerdan a las obligaciones de gobernanza de datos y evaluación de sesgos impuestas en la Unión Europea a los sistemas de alto riesgo, pero traduce estas obligaciones en deberes directamente oponibles al empleador desde la óptica del Derecho antidiscriminatorio y de la regulación municipal.

Junto a la experiencia neoyorquina, cabe mencionar la Artificial Intelligence Video Interview Act del Estado de Illinois⁷⁶, que re-

⁷⁶ <https://www.workforcebulletin.com/illinois-prohibits-discriminatory-artificial-intelligence-in-employment-decisions>

gula el uso de sistemas de IA para analizar entrevistas en vídeo de personas candidatas. Esta norma no impone auditorías de sesgo en sentido estricto, pero establece tres obligaciones básicas antes de recabar y analizar entrevistas mediante IA: a) notificar a cada solicitante que se empleará inteligencia artificial para evaluar su entrevista; b) proporcionar información comprensible sobre el funcionamiento general del sistema y las características que se valorarán; y c) obtener el consentimiento expreso de la persona candidata para ser evaluada mediante IA. Además, cuando el empleador se basa exclusivamente en el análisis automatizado para decidir qué candidaturas pasan a entrevistas presenciales, debe recopilar y remitir a las autoridades información desagregada sobre la raza y el sexo de las personas entrevistadas y seleccionadas, con la finalidad de detectar posibles patrones de discriminación.

Desde la perspectiva de la igualdad, la relevancia de esta regulación reside en dos elementos. En primer lugar, reconoce que el uso de IA en entrevistas de trabajo plantea riesgos específicos de discriminación –incluida la de género– que exigen un plus de transparencia y control, reforzando el derecho de la persona candidata a comprender los criterios generales de evaluación. En segundo lugar, al exigir la recogida y reporte de datos desagregados por sexo y raza, la norma asume la necesidad de disponer de información sensible para detectar posibles sesgos de selección, en línea con los debates doctrinales sobre la compatibilidad entre protección de datos y lucha contra la discriminación algorítmica⁷⁷.

En conjunto, estas experiencias subnacionales muestran una aproximación regulatoria que, sin configurar un estatuto laboral de la IA, introduce obligaciones claras de auditoría, transparencia y recogida de datos desagregados cuando la IA interviene en decisiones de empleo. Su interés comparado para el Derecho del

⁷⁷ <https://www.crowell.com/en/insights/client-alerts/artificial-intelligence-in-employment-update-illinois-requires-notice-and-prohibits-discriminatory-impact-in-use-of-ai>

Trabajo europeo radica, sobre todo, en tres aspectos: la normalización de las auditorías de impacto dispar como herramienta estándar; la traslación de obligaciones directamente a los empleadores que usan herramientas desarrolladas por terceros; y el reconocimiento de que la supervisión pública de estos sistemas puede articularse también desde ámbitos regulatorios distintos del puramente laboral, pero con efectos intensos sobre la igualdad en el trabajo.

5.2. INTERVENCIÓN DE AUTORIDADES LABORALES Y ANTIDISCRIMINATORIAS EN EL USO DE IA CON ENFOQUE DE IGUALDAD

Más allá de la creación de normas específicas, la consolidación de un “derecho del trabajo algorítmico” depende en buena medida del modo en que las autoridades encargadas de garantizar la igualdad y los derechos laborales incorporan la dimensión algorítmica a sus prácticas de supervisión y sanción. En este terreno, aunque todavía no existe un patrón uniforme, pueden identificarse algunas experiencias pioneras que ilustran cómo inspecciones de trabajo y autoridades antidiscriminatorias comienzan a abordar la IA en clave de igualdad de género y no discriminación.

En el plano estadounidense, la Equal Employment Opportunity Commission (EEOC) ha emitido en 2023 una guía específica⁷⁸ sobre el uso de software, algoritmos e inteligencia artificial en los procedimientos de selección regulados por el Title VII. La EEOC parte de la premisa de que las herramientas de IA empleadas para contratar, promocionar o despedir pueden constituir “procedimientos de selección” a los efectos del derecho antidiscriminatorio federal y, por tanto, quedar sometidas a las reglas sobre disparate impact. La guía aclara que el hecho de que el algoritmo haya sido desarrollado por un proveedor externo no exime al empleador de responsabilidad cuando el uso de la herramienta produce

⁷⁸ <https://www.mayerbrown.com/en/insights/publications/2023/07/eec-issues-title-vii-guidance-on-employer-use-of-ai-other-algorithmic-decisionmaking-tools>

efectos adversos desproporcionados para mujeres u otros grupos protegidos, y subraya la obligación empresarial de evaluar periódicamente si sus herramientas generan impactos diferenciales.

Desde la óptica de la igualdad de género, la intervención de la EEOC es relevante por varias razones. En primer lugar, reafirma que las categorías protegidas tradicionales –incluyendo el sexo– siguen siendo plenamente aplicables en contextos de toma de decisiones automatizada, de modo que un algoritmo que excluya sistemáticamente a mujeres de determinados puestos puede vulnerar el Title VII aunque el código no incluya la variable “sexo” de forma explícita. En segundo lugar, la guía impulsa un enfoque proactivo de gestión del riesgo discriminatorio, al exigir a las empresas que “conozcan” el comportamiento de sus herramientas y que revisen sus decisiones cuando detecten impactos desproporcionados, aproximándose a la lógica de los sistemas de gestión de riesgos y gobernanza de datos que el AI Act impone en Europa.

En el ámbito europeo, la intervención directa de autoridades con competencias laborales o de igualdad sobre sistemas de IA se encuentra en una fase más embrionaria, pero empiezan a vislumbrarse líneas de actuación. De un lado, algunas autoridades de protección de datos han dictado guías sobre tratamientos que incluyen IA, subrayando la necesidad de incorporar evaluaciones de impacto específicas y de analizar los riesgos de discriminación, incluida la de género, en los procesos de auditoría⁷⁹. De otro, las inspecciones de trabajo están intensificando el control sobre las obligaciones empresariales de igualdad en sentido amplio, lo que abre la puerta a integrar progresivamente el dimensionamiento algorítmico dentro del diagnóstico y la verificación de los planes de igualdad.

En el caso español, por ejemplo, la Inspección de Trabajo y Seguridad Social ha desarrollado campañas específicas para verifi-

⁷⁹ CRISTOFOLINI CH. “Navigating the impact of AI systems in the workplace: strengths and loopholes of the EU AI Act from a labor perspective”, *Italian Labour Law e-Journal*, Issue 1, Vol. 17 (2024).

car el cumplimiento de las obligaciones en materia de planes de igualdad, registros retributivos y auditorías salariales, exigiendo documentación detallada sobre brechas salariales, segregación ocupacional y medidas de corrección. Aunque estas actuaciones no se dirigen todavía de forma expresa a la IA, el marco procedimental –diagnóstico cuantitativo, análisis de brechas, exigencia de medidas correctoras y posibilidad de sanción– resulta fácilmente adaptable a la supervisión de sistemas algorítmicos cuando estos inciden en la organización del trabajo, la promoción o la remuneración. En particular, la obligación de justificar brechas salariales superiores al 25% y de introducir medidas correctoras en los planes de igualdad⁸⁰ puede servir como palanca para exigir a las empresas que auditen también el impacto de sus herramientas automatizadas en la estructura retributiva y en la asignación de complementos, especialmente allí donde se utilizan algoritmos para evaluar el rendimiento o distribuir incentivos.

Del mismo modo, la publicación por parte del Ministerio de Trabajo de herramientas oficiales de valoración de puestos con perspectiva de igualdad, destinadas a facilitar la realización de auditorías salariales, anticipa un posible escenario en el que las propias administraciones laborales desarrollen o certifiquen instrumentos de análisis algorítmico que incorporen la dimensión de género. Aunque, de momento, estas herramientas se centran en la valoración de puestos a efectos de igualdad retributiva, su lógica –sistemas estandarizados, parámetros verificables, documentación obligatoria– podría extenderse a la evaluación de sistemas de IA de alto riesgo utilizados en la gestión de personas, en estrecha coordinación con las autoridades de protección de datos y de igualdad.

Por último, cabe mencionar que algunas autoridades europeas de igualdad y organismos de supervisión han comenzado a explorar metodologías para detectar anuncios de empleo discriminatorios

⁸⁰ ROMERO RODENAS M.J., *Planes de igualdad, régimen jurídico, negociación y práctica aplicada*, Bomarzo, Albacete, 2026.

y patrones de segmentación algorítmica en redes sociales y plataformas, utilizando técnicas de IA para identificar sesgos en la difusión de ofertas. Estas experiencias, aunque centradas en la fase de oferta de empleo, muestran que la intervención pública puede servirse también de herramientas algorítmicas para vigilar a los algoritmos, reforzando la tutela del derecho a la igualdad de trato en las fases preliminares de la relación laboral.

En conjunto, las experiencias comparadas analizadas revelan un doble movimiento. Por un lado, algunos ordenamientos comienzan a imponer obligaciones específicas a los empleadores que utilizan IA en la selección de personal, con especial atención a la realización de auditorías de sesgo y a la transparencia frente a las personas candidatas. Por otro, las autoridades laborales y antidiscriminatorias, en particular la EEOC en Estados Unidos y organismos de igualdad en Europa, están adaptando sus herramientas clásicas de supervisión –evaluaciones de impacto, control de brechas, exigencia de justificación objetiva– a los desafíos de la toma de decisiones automatizada. La convergencia entre ambas tendencias apunta hacia un modelo en el que la igualdad en el empleo deja de ser un mero límite externo a la IA para convertirse en un criterio estructural de su diseño, despliegue y control, tanto en la fase de acceso al trabajo como en el desarrollo de la relación laboral.

6. GOBERNANZA FEMINISTA DE LA IA EN LA EMPRESA: MÁS ALLÁ DEL CUMPLIMIENTO TÉCNICO. PROPUESTAS

Ya hemos señalado que la incorporación de la inteligencia artificial en la empresa está siendo guiada, en gran medida, por criterios de eficiencia, reducción de costes y ventaja competitiva, sin que la igualdad de género forme parte explícita de los objetivos de diseño y despliegue. Ello provoca que incluso sistemas formalmente conformes con el Reglamento de IA continúen reproduciendo estructuras de desigualdad, porque han sido concebidos desde una lógica empresarial neutra solo en apariencia. Frente a este enfoque, resulta necesario construir una gobernanza feminista de la IA que sitúe la igualdad sustantiva entre mujeres y hombres como criterio rector del ciclo completo del sistema, desde la definición del caso de uso hasta la supervisión posterior a la comercialización.

Una gobernanza feminista implica, en primer lugar, revisar quién decide qué problemas se automatizan y con qué objetivos. No es neutro que la IA se utilice prioritariamente para optimizar la productividad o reducir la “rotación”, si ello conlleva penalizar trayectorias laborales discontinuas asociadas a los cuidados o prácticas de conciliación. En segundo lugar, exige incorporar de manera estable la perspectiva de género en los órganos internos de decisión sobre datos y algoritmos: comités de ética, unidades de cumplimiento o departamentos de recursos humanos no pueden seguir siendo espacios homogéneos sin presencia de expertos en igualdad. Esta gobernanza requiere mecanismos de rendición de cuentas que vayan más allá de las métricas de rendimiento técnico e incluyan indicadores de impacto de género, revisados pe-

riódicamente con participación de la representación legal de las personas trabajadoras.

6.1. INCORPORAR LA IGUALDAD DE GÉNERO EN LA POLÍTICA ALGORÍTMICA CORPORATIVA: PROPUESTAS

Exigir que las empresas que utilicen sistemas de IA de alto riesgo en el ámbito laboral aprueben una “política algorítmica corporativa con perspectiva de género”, integrada en sus políticas de cumplimiento normativo y planes de igualdad.

Esta política debería:

- Identificar los procesos laborales donde se usa IA (selección, promoción, evaluación, horarios, etc.) y describir sus posibles impactos diferenciados por sexo.
- Establecer objetivos específicos de igualdad (por ejemplo, no incremento de brechas salariales, de promoción o de estabilidad contractual tras la implantación de IA).
- Definir indicadores cuantitativos y cualitativos de impacto de género para cada sistema (tasas de selección, puntuaciones medias, tasas de error, etc., desagregadas por sexo).

La aprobación de esta política podría configurarse como obligación legal para empresas que superen determinados umbrales de plantilla o volumen de negocio, o bien como requisito en los planes de igualdad de las empresas que usen IA de alto riesgo.

6.2. REFORZAR LA COMPOSICIÓN Y FUNCIONES DE LOS ÓRGANOS INTERNOS DE DECISIÓN: PROPUESTAS

Exigir que los comités de ética, comités de empresa o comités de datos o estructuras equivalentes incluyan al menos una persona experta en igualdad de género y no discriminación, con voz y voto en las decisiones sobre diseño, compra o despliegue de sistemas algorítmicos.

Introducir, en la normativa de igualdad y en la de IA, la obligación de que estos órganos:

- Revisar los análisis de riesgo de cada sistema, con foco específico en discriminación directa e indirecta por razón de sexo.
- Validar los criterios de éxito de los modelos (por ejemplo, “disponibilidad”, “estabilidad”, “productividad”) para detectar si incorporan estereotipos de género.
- Aprobar planes de mitigación cuando se observen impactos desproporcionados sobre las mujeres, e incorporar este tipo de medidas en el Plan de Igualdad de la empresa.

Esta institucionalización de la perspectiva de género responde a la idea, ya presente en nuestro estudio, de que la IA es un nuevo dispositivo de poder empresarial y debe someterse a controles semejantes a otros poderes organizativos.

6.3. TRANSPARENCIA INTERNA Y RENDICIÓN DE CUENTAS CON ENFOQUE DE GÉNERO: PROPUESTAS

Establecer legalmente la obligación de elaborar un “informe anual de impacto de la IA en la igualdad de género” que bien puede venir su exigencia en el Plan de Igualdad en las empresas que utilicen sistemas de alto riesgo, que:

- Se integre en la información de los planes de igualdad y en el registro retributivo.
- Incluya resultados desagregados por sexo sobre decisiones automatizadas (selección, promociones, evaluaciones, rescisión de contratos, asignación de horarios).
- Sea accesible a la representación legal de las personas trabajadoras y, en su caso, a las autoridades de igualdad y de la Inspección de Trabajo.

Este informe reforzaría la rendición de cuentas y dotaría de contenido material a las obligaciones de transparencia del AI Act, hoy formuladas en términos fundamentalmente técnicos. Nuestro análisis subraya que el Reglamento de IA se centra en la corrección técnica del sesgo, sin garantizar por sí solo la igualdad sustantiva, lo que obliga a complementar su cumplimiento con políticas internas que incorporen la perspectiva de género.

7. IGUALDAD Y NEGOCIACIÓN COLECTIVA ANTE LA IA: CLÁUSULAS DE NO DISCRIMINACIÓN ALGORÍTMICA. PROPUESTAS

La negociación colectiva constituye un espacio privilegiado para traducir el marco europeo y estatal de protección frente a la discriminación algorítmica en garantías concretas dentro de la empresa. Si tradicionalmente los convenios se han ocupado de regular salarios, jornada o clasificación profesional, hoy resulta imprescindible que incorporen cláusulas específicas sobre uso de algoritmos y sistemas de IA en selección, promoción, evaluación del desempeño y organización del tiempo de trabajo, con especial atención a los impactos de género. Ello enlaza con la función estructural de la negociación colectiva como instrumento de igualdad sustantiva, reconocida tanto por el derecho interno como por el acervo europeo en materia de no discriminación⁸¹.

Estas cláusulas pueden articular, al menos, tres tipos de obligaciones. En primer lugar, un deber de información reforzada hacia la representación legal de las personas trabajadoras sobre la existencia de sistemas automatizados, su finalidad y las variables más relevantes que intervienen en la toma de decisiones. En segundo lugar, la creación de comisiones paritarias de seguimiento algorítmico, con acceso a auditorías internas y resultados desagregados por sexo u otras categorías protegidas, encargadas de identificar

⁸¹ El V Acuerdo por el Empleo y la Negociación Colectiva (AENC) contiene un capítulo sobre la Inteligencia Artificial (IA), una cláusula pionera en el mundo de la negociación colectiva, prácticamente sin precedentes en Europa, que ejemplifica cómo el diálogo social puede situar la regulación muy cerca de la realidad tecnológica, vid. Informe elaborado por UGT, noviembre 2023 <https://servicioestudiosugt.com/recomendaciones-para-la-negociacion-colectiva-de-la-ia/>

impactos adversos y proponer medidas correctoras. En tercer lugar, el reconocimiento convencional de un derecho individual a la revisión humana cualificada de decisiones automatizadas relevantes, vinculado a un procedimiento específico y ágil de reclamación, de modo que el ejercicio del derecho a la explicación previsto en el AI Act se inserte funcionalmente en el sistema de garantías laborales y antidiscriminatorias.

En páginas anteriores, destacábamos la centralidad de la negociación colectiva como espacio para pactar mecanismos de transparencia reforzada, protocolos de evaluación de sesgos y cláusulas de no discriminación algorítmica que vayan más allá del mínimo normativo. Es más la articulación de derechos colectivos de información y consulta sobre algoritmos se apoya en la tradición del art. 64 ET y en las obligaciones de transparencia sobre decisiones automatizadas presentes en el RGPD (arts. 13 y 22) y la LO 3/2018. Por tanto, la negociación colectiva es un cauce idóneo para traducir el marco europeo y estatal en garantías operativas frente a la discriminación algorítmica⁸².

7.1. CLÁUSULAS TIPO SOBRE INFORMACIÓN Y CONSULTA: PROPUESTAS

Incluir en los convenios colectivos cláusulas que reconozcan a la representación legal de las personas trabajadoras el derecho a:

- Recibir información previa y detallada sobre cualquier sistema de IA o algoritmo que se utilice para tomar decisiones que afecten a la contratación, condiciones de trabajo, promoción o extinción.
- Ser consultada con carácter previo a la implantación o modificación sustancial de estos sistemas, equiparando su relevancia a la de decisiones organizativas clásicas (art. 64 ET).

⁸² Acuerdo Marco Europeo de los interlocutores Sociales sobre digitalización, CCOO, marzo 2021. <https://www.ccoo.es/3ec9e3ddff84034c1a796cb52ac84c09000001.pdf>

La cláusula debería especificar el contenido mínimo de la información: finalidad del sistema, tipos de datos utilizados (incluida posible presencia de proxies), variables de decisión más influyentes y mecanismos de supervisión humana.

7.2. COMISIONES PARITARIAS DE SEGUIMIENTO ALGORÍTMICO: PROPUESTAS

Prever en convenio colectivo la creación de “comisiones paritarias de seguimiento de sistemas algorítmicos”, con funciones de:

- Acceso a resultados desagregados por sexo y otros factores relevantes (edad, tipo de contrato, tiempo parcial, etc.).
- Participación en el diseño y seguimiento de auditorías internas o externas de IA.
- Formulación de recomendaciones vinculantes cuando se detecten impactos adversos de género, con obligación de la empresa de adoptar medidas correctoras en un plazo concreto.

Estas comisiones podrían integrarse en las actuales comisiones de elaboración y seguimiento del plan de igualdad, evitando la proliferación desordenada de órganos y reforzando la coherencia del sistema.

7.3. DERECHOS INDIVIDUALES DE REVISIÓN Y EXPLICABILIDAD LABORAL: PROPUESTAS

Incorporar en los convenios colectivos y acuerdos de empresa:

- Un derecho individual a solicitar revisión humana de las decisiones relevantes que se apoyen en sistemas automatizados (selección, promociones, evaluaciones negativas, decisiones disciplinarias), con plazo y procedimiento definidos.
- Un derecho a recibir una explicación comprensible que identifique los criterios determinantes de la decisión y su relación con los datos de la persona afectada, adaptando las obliga-

ciones de explicabilidad técnica a las exigencias probatorias del Derecho del Trabajo.

Estas cláusulas conectarían el derecho a no ser sometida exclusivamente a decisiones automatizadas del art. 22 RGPD con las garantías laborales clásicas (impugnación del despido, tutela de derechos fundamentales, etc.), mitigando la “opacidad algorítmica” que dificulta la acreditación de la discriminación.

8. IA, CUIDADOS Y ORGANIZACIÓN DEL TIEMPO DE TRABAJO: RIESGOS DE DISCRIMINACIÓN INDIRECTA. PROPUESTAS

Uno de los puntos ciegos de la regulación actual es el impacto de los sistemas de IA sobre la organización del tiempo de trabajo y, en particular, sobre quienes asumen mayor carga de cuidados, que siguen siendo mayoritariamente mujeres. Algoritmos destinados a optimizar turnos, asignar horarios o predecir “disponibilidad” pueden operar con criterios aparentemente neutros –maximización de horas en franjas de mayor demanda, flexibilidad absoluta, reducción de absentismo– que, en la práctica, penalizan a quienes solicitan adaptaciones horarias, reducciones de jornada o teletrabajo para conciliación. Esta penalización puede traducirse en menores oportunidades de promoción, peores evaluaciones del desempeño o mayor probabilidad de extinción contractual, configurando supuestos de discriminación indirecta por razón de sexo.

Para anticipar y corregir estos riesgos, resulta imprescindible introducir la perspectiva de género en el diseño de los modelos de planificación y en los indicadores de éxito que se utilizan. No es admisible, desde el prisma del Derecho del Trabajo, que la “disponibilidad total” se configure como criterio central de empleabilidad o mérito sin valorar su impacto desproporcionado sobre las mujeres, ni que se ignoren las obligaciones empresariales de adaptar la organización del tiempo de trabajo a las necesidades de conciliación reconocidas normativamente. La evaluación de impacto en igualdad debería convertirse en requisito inexcusable antes de desplegar sistemas de IA que afecten a turnos, horarios o asignación de tareas, combinando la metodología de las evalua-

ciones de impacto algorítmico con la experiencia acumulada en los planes de igualdad.

Nuestro análisis muestra cómo los sistemas de IA pueden penalizar trayectorias laborales marcadas por interrupciones o jornadas reducidas asociadas a los cuidados, convirtiendo la desigualdad histórica en una característica estructural del futuro del trabajo. Se evidencia la persistente brecha de género en responsabilidad de cuidados y en condiciones de empleo (tiempo parcial, salarios, estabilidad), lo que multiplica el riesgo de discriminación indirecta cuando se automatizan decisiones sobre organización del trabajo.

8.1. EVALUACIONES DE IMPACTO DE GÉNERO ANTES DE AUTOMATIZAR LA ORGANIZACIÓN DEL TIEMPO DE TRABAJO: PROPUESTAS

Imponer la realización obligatoria de una “evaluación de impacto de género” previa a la implantación de sistemas de IA destinados a:

- Planificar turnos y horarios.
- Asignar carga de trabajo.
- Medir la “disponibilidad” o “flexibilidad” como criterios de mérito.

Esta evaluación debería:

- Analizar la situación de partida (distribución de trabajadores y trabajadoras por tipo de jornada, uso de medidas de conciliación, etc.).
- Simular el comportamiento del sistema sobre distintos perfiles (personas con reducción de jornada, con adaptaciones por cuidado, con teletrabajo parcial).
- Identificar posibles impactos desproporcionados y prever medidas correctoras (ponderación menos penalizadora de determinadas ausencias, límites a la utilización de determinados indicadores, etc.).

8.2. REVISIÓN DE CRITERIOS DE “IDONEIDAD” Y “RENDIMIENTO”: PROPUESTAS

Incorporar en la normativa y en la negociación colectiva una obligación de revisar los criterios de rendimiento y disponibilidad utilizados por los algoritmos, de manera que:

- No se puedan utilizar directamente, como señales negativas, el uso legítimo de derechos de conciliación (reducciones de jornada, adaptaciones horarias, permisos vinculados con los cuidados y conciliación).
- Se consideren “datos protegidos” a efectos internos ciertas variables estrechamente asociadas a la maternidad y a los cuidados, de forma que su utilización requiera una justificación reforzada y se someta a revisión sindical.

Se trata de una propuesta anclada en la lógica de la discriminación indirecta: aun cuando el criterio aplicado sea formalmente neutro, su efecto práctico puede recaer de manera desproporcionada sobre las mujeres si no se corrige.

8.3. INTEGRAR LA IA EN LOS PLANES DE IGUALDAD Y EN LOS DIAGNÓSTICOS DE BRECHA: PROPUESTAS

Incorporar, como contenido mínimo de los planes de igualdad:

- Un eje específico sobre “uso de sistemas de IA y algoritmos en la empresa”, con inventario de sistemas utilizados y análisis de su impacto sobre las brechas de género existentes.
- Indicadores concretos que relacionen decisiones automatizadas con diferencias por sexo en contratación, promoción, formación, retribución y estabilidad contractual.

Ello permitiría utilizar los instrumentos ya conocidos (diagnóstico de situación, medidas, seguimiento) para incorporar los nuevos riesgos algorítmicos, evitando crear un ámbito paralelo desconectado de la política de igualdad.

Nuestro análisis pone de relieve la persistencia de la infrarrepresentación de mujeres en puestos de responsabilidad y en el sector tecnológico, así como la concentración en contratos a tiempo parcial y menos estables, factores que aumentan la vulnerabilidad frente a criterios automatizados de disponibilidad y continuidad.

9. AUTORIDADES DE CONTROL Y PERSPECTIVA DE GÉNERO: HACIA UNA SUPERVISIÓN PROACTIVA DE LA IA LABORAL

Las obligaciones preventivas del Reglamento de Inteligencia Artificial sólo pueden desplegar plenamente su potencial si las autoridades de control asumen un papel activo en la detección y corrección de los sesgos que atraviesan la vida laboral. La supervisión de la IA utilizada en la selección, gestión y evaluación de las personas trabajadoras ya no puede limitarse a verificar el cumplimiento formal de requisitos técnicos o de protección de datos, sino que exige incorporar de manera sistemática una perspectiva de género capaz de identificar impactos diferenciales sobre mujeres y hombres, así como sobre otros colectivos vulnerables. Este epígrafe examina hasta qué punto los actuales mandatos de las autoridades nacionales –de igualdad, de protección de datos, de inspección de trabajo o de vigilancia del mercado– permiten una supervisión proactiva de los sistemas de IA laboral y qué transformaciones normativas y organizativas serían necesarias para que estos organismos se conviertan en piezas clave de la garantía efectiva del derecho a la igualdad y a la no discriminación en el trabajo mediado por algoritmos.

Conocemos que el despliegue del AI Act ha dado lugar a una nueva arquitectura institucional europea en materia de inteligencia artificial, en la que destacan la Oficina Europea de IA, el grupo de expertos científicos independientes y las autoridades nacionales de vigilancia del mercado. Sin embargo, en nuestra opinión, la integración efectiva de la perspectiva de género en la actuación de estas instancias está lejos de estar garantizada: ni el diseño de procedimientos de supervisión ni las metodologías de evaluación de riesgos incluyen, de forma sistemática, indicadores de impac-

to sobre la igualdad entre mujeres y hombres. Existe el riesgo de que la supervisión se concentre en aspectos puramente técnicos de seguridad y rendimiento, relegando a un segundo plano las dimensiones relacionadas con la discriminación algorítmica.

Para evitarlo, resulta imprescindible que las autoridades de control incorporen, en sus guías, criterios y prácticas de supervisión, obligaciones específicas de análisis de género. Ello pasa por exigir que las evaluaciones de conformidad y las auditorías de sistemas de alto riesgo incluyan métricas desagregadas por sexo y otros factores de discriminación, así como por priorizar en su actuación los sectores donde confluyen mayor riesgo algorítmico y brechas de género preexistentes (empleo, protección social, crédito). Además, debe reforzarse la cooperación entre autoridades de vigilancia del mercado, organismos de igualdad y autoridades de protección de datos, de modo que la supervisión de la IA laboral no se fragmente en silos institucionales que dificulten una comprensión integrada de los riesgos para la igualdad.

Con anterioridad hemos señalado que el conflicto entre secreto empresarial y transparencia técnica no puede resolverse sacrificando el derecho a la igualdad, por lo que las autoridades deberán utilizar al máximo sus facultades de requerimiento de información cuando estén en juego posibles sesgos de género. El AI Act crea una arquitectura institucional potente, pero la integración de la perspectiva de género en su funcionamiento no está garantizada y exige decisiones políticas y técnicas adicionales.

9.1. MANDATOS EXPLÍCITOS DE IGUALDAD PARA AUTORIDADES DE IA, DE DATOS Y DE IGUALDAD: PROPUESTAS

Modificar o interpretar la normativa nacional para:

- Incluir, entre las funciones de la autoridad nacional competente en IA, la obligación expresa de vigilar los riesgos de discriminación por razón de sexo en los sistemas de alto riesgo utilizados en el ámbito laboral.

- Establecer mecanismos estables de cooperación entre la autoridad de IA, la autoridad de protección de datos y los organismos de igualdad (por ejemplo, el Instituto de las Mujeres), de modo que puedan compartir información y coordinar investigaciones cuando se detecten posibles sesgos de género.

9.2. PROTOCOLOS DE SUPERVISIÓN CON INDICADORES DE GÉNERO: PROPUESTAS

Desarrollar guías y protocolos de supervisión que:

- Exigir a los proveedores y usuarios de IA la presentación de métricas de rendimiento desagregadas por sexo (y, cuando proceda, por otras variables relevantes) como parte de la documentación que acompaña a los sistemas de alto riesgo.
- Incluir indicadores mínimos para identificar riesgos de discriminación (por ejemplo, diferencias significativas en tasas de selección, puntuaciones medias o tasas de error entre mujeres y hombres).
- Establecer umbrales a partir de los cuales deba iniciarse una evaluación más profunda o una auditoría específica.

En este sentido es muy importante la guía de la AEPD sobre auditorías de IA⁸³, que ya recomienda comparativas de rendimiento por características sensibles y métricas de equidad, incorporando de forma más clara la perspectiva de género.

9.3. USO ESTRATÉGICO DEL GRUPO DE EXPERTOS CIENTÍFICOS: PROPUESTAS

Priorizar, en las solicitudes de asistencia al grupo de expertos científicos independientes previstas en el art. 14 del AI Act, aquellos casos en que:

⁸³ <https://www.aepd.es/guias/requisitos-auditorias-tratamientos-incluyan-ia.pdf>

- ♦ El sistema de IA se utilice para selección, promoción o evaluación en sectores con alta feminización o alta brecha de género.
- ♦ Se hayan detectado, por denuncias o indicadores internos, patrones sistemáticos de desventaja para las mujeres (por ejemplo, tasas de selección significativamente inferiores o concentraciones de mujeres en los percentiles inferiores de puntuación).

Además, se podría promover que en las metodologías de evaluación que aplique este grupo se incluya de manera obligatoria la revisión de sesgos de género y de otros motivos de discriminación interseccionales.

Nuestro estudio destaca que el Reglamento de ejecución (UE) 2025/454 exige equilibrio de género en la composición del grupo de expertos, pero ello no basta si sus tareas y métodos no incorporan de forma explícita la evaluación de impactos sobre la igualdad.

10. CONCLUSIONES

En este apartado se sintetizan, desde una perspectiva jurídico-laboral y de género, las principales aportaciones realizadas sobre la interacción entre la inteligencia artificial y el derecho a la igualdad y no discriminación, valoramos sus implicaciones para la práctica judicial y para la acción de las autoridades de control, y se señalan algunas líneas de reforma e investigación futura que podrían contribuir a una protección más efectiva frente a la discriminación algorítmica en el ámbito del Derecho del Trabajo y de la Seguridad Social.

10.1. LA IA EN EL TRABAJO NO ES NEUTRA: REPRODUCE Y PUEDE AMPLIFICAR LAS DESIGUALDADES DE GÉNERO

La investigación confirma que la inteligencia artificial aplicada al trabajo no es una tecnología neutra, ni en términos sociales ni en términos de género. Los sistemas algorítmicos que intervienen en selección, promoción, evaluación del desempeño u organización del tiempo de trabajo aprenden de datos históricos profundamente marcados por la segregación ocupacional, el techo de cristal, la penalización de la maternidad y la desigual distribución de los cuidados. En estas condiciones, la IA tiende a convertir en criterio de futuro lo que fueron decisiones discriminatorias del pasado, consolidando la desigualdad bajo una apariencia de objetividad técnica.

La opacidad de muchos modelos –especialmente los basados en aprendizaje profundo– dificulta además la identificación y la

prueba de la discriminación, trasladando los sesgos a estructuras técnicas menos visibles y más difíciles de cuestionar. Todo ello obliga a reafirmar un principio de partida: la digitalización de las relaciones laborales, y en particular la adopción de sistemas de IA, sólo es legítima si se somete a los límites derivados del derecho fundamental a la igualdad y a la prohibición de discriminación por razón de sexo, tal como han sido configurados por el Derecho de la Unión Europea y por el ordenamiento español.

10.2. EL MARCO NORMATIVO VIGENTE ES NECESARIO PERO INSUFICIENTE FRENTE A LA DISCRIMINACIÓN ALGORÍTMICA DE GÉNERO

El análisis del Reglamento europeo de IA, del RGPD, de la LO 3/2018, de la normativa anti-discriminatoria y de las leyes de igualdad muestra un avance significativo en la identificación de riesgos y en la imposición de salvaguardas técnicas (gestión de riesgos, gobernanza de datos, supervisión humana, documentación y transparencia). Sin embargo, este entramado normativo sigue siendo insuficiente para garantizar una igualdad sustantiva entre mujeres y hombres en contextos de decisión automatizada.

Por un lado, el AI Act está construido fundamentalmente como una norma de seguridad de productos y de gestión de riesgos tecnológicos, no como un instrumento central de política de igualdad. Sus obligaciones se orientan a corregir “sesgos indebidos” en los datos, pero no traducen en términos operativos las categorías jurídicas de discriminación directa, indirecta o por asociación, ni las exigencias de acción positiva que caracterizan el derecho de igualdad de trato. Por otro lado, el RGPD y la LO 3/2018 proporcionan herramientas importantes (derecho a no ser objeto de decisiones automatizadas, evaluaciones de impacto, derechos de información), pero su aplicación al ámbito laboral con enfoque de género aún es embrionaria y carece de pautas específicas sobre cómo valorar los efectos diferenciados por sexo de los perfiles algorítmicos.

La consecuencia es una brecha entre el cumplimiento formal de las obligaciones técnicas y la efectividad real del derecho a la no discriminación: un sistema puede superar auditorías de datos y pruebas de rendimiento y seguir produciendo resultados sistemáticamente desfavorables para las mujeres si no se analiza de forma explícita su impacto de género.

10.3. LA GOBERNANZA DE LA IA DEBE FEMINIZARSE: IGUALDAD SUSTANTIVA COMO CRITERIO RECTOR DE DISEÑO Y USO

La investigación pone de relieve que buena parte de los riesgos de discriminación no se sitúan sólo en el nivel de los datos, sino en decisiones de diseño y de gobernanza: qué problemas se automatizan, con qué objetivos, qué variables se consideran relevantes, quién participa en estas decisiones y con qué formación en igualdad. La infrarrepresentación de mujeres y de perfiles con competencias en género en los equipos que diseñan, adquieren y supervisan sistemas de IA constituye un factor estructural de vulnerabilidad.

De ahí la necesidad de plantear una “gobernanza feminista de la IA” en la empresa, que incorpore la perspectiva de género en toda la cadena de decisión: políticas algorítmicas corporativas vinculadas a los planes de igualdad, órganos internos con presencia de personas expertas en igualdad, informes periódicos de impacto de la IA en la brecha de género, e integración de indicadores de igualdad en los sistemas de gestión de riesgos y de auditoría. Esta reorientación no puede quedar en la autorregulación voluntaria, sino que requiere anclaje normativo, tanto en el desarrollo interno del AI Act como en la legislación laboral e igualitaria española.

10.4. LA NEGOCIACIÓN COLECTIVA ES UN INSTRUMENTO CLAVE PARA CONSTRUIR GARANTÍAS ESPECÍFICAS FRENTE A LA DISCRIMINACIÓN ALGORÍTMICA

El estudio confirma el potencial de la autonomía colectiva para adaptar el marco jurídico general a las particularidades sectoria-

les y empresariales de la IA. Igual que en otros momentos del desarrollo del Derecho del Trabajo la negociación colectiva ha sido el espacio donde se han concretado avances en igualdad retributiva, conciliación o prevención del acoso, hoy debe asumir un papel central en la regulación de los sistemas algorítmicos que afectan a las condiciones de trabajo.

En particular, los convenios colectivo y acuerdos colectivos pueden:

- ♦ Reconocer derechos de información y consulta reforzados sobre la implantación y modificación de sistemas de IA.
- ♦ Establecer comisiones paritarias de seguimiento algorítmico con competencias de acceso a auditorías y datos desagregados por sexo.
- ♦ Pactar cláusulas de transparencia y explicabilidad adaptadas al lenguaje jurídico-laboral.
- ♦ Prever derechos individuales de revisión humana de decisiones automatizadas y de impugnación específica cuando se alegue discriminación por razón de sexo.

La negociación colectiva es clave, se revela así como un espacio imprescindible para materializar en la práctica empresarial el mandato de igualdad sustantiva, evitando que los algoritmos se conviertan en un nuevo ámbito opaco de ejercicio del poder empresarial.

10.5. IA, CUIDADOS Y ORGANIZACIÓN DEL TIEMPO DE TRABAJO: EL NÚCLEO DE LA DISCRIMINACIÓN INDIRECTA

Uno de los hallazgos más relevantes de la investigación es la centralidad de la organización del tiempo de trabajo y de los cuidados en la configuración de la discriminación algorítmica de género. Sistemas diseñados para optimizar turnos, medir disponibilidad o predecir riesgos de rotación pueden penalizar sistemáticamente a

quienes ejercen derechos de conciliación, acumulan reducciones de jornada o presentan trayectorias laborales discontinuas, esto es, a las mujeres en su mayoría.

La automatización, si no se corrige, corre el riesgo de reforzar un modelo de persona trabajadora ideal permanentemente disponible, sin responsabilidades externas, incompatible con la vida de cuidados. Esto convierte la IA en un vector particularmente potente de discriminación indirecta por razón de sexo, al transformar en requisito técnico lo que el Derecho del Trabajo lleva décadas intentando corregir a través de normas de conciliación, corresponsabilidad y protección frente al despido discriminatorio.

Frente a ello, el estudio propone integrar evaluaciones de impacto de género en la implantación de sistemas de IA que afecten al tiempo de trabajo, revisar críticamente los criterios de rendimiento y disponibilidad usados por los modelos y conectar la política algorítmica con los planes de igualdad y los diagnósticos de brecha existentes en las empresas.

10.6. LAS AUTORIDADES DE CONTROL DEBEN ASUMIR UN MANDATO EXPLÍCITO DE IGUALDAD EN LA SUPERVISIÓN DE LA IA

La nueva arquitectura institucional prevista por el AI Act (Oficina Europea de IA, grupo de expertos científicos, autoridades nacionales de vigilancia del mercado) ofrece una oportunidad para reforzar la tutela frente a la discriminación algorítmica, pero sólo si la perspectiva de género se incorpora explícitamente a su mandato y a sus metodologías de trabajo.

El estudio muestra que, en ausencia de directrices claras, la supervisión corre el riesgo de centrarse en aspectos puramente técnicos (seguridad, robustez, documentación) y dejar en un segundo plano los impactos sobre los derechos fundamentales y, en particular, sobre la igualdad entre mujeres y hombres. Por ello, se propone:

- Integrar en los protocolos de supervisión indicadores mínimos de impacto de género (tasas de error y selección desagregadas, diferencias significativas en resultados por sexo, etc.).
- Priorizar las investigaciones y auditorías en sectores y usos de IA donde la combinación de riesgo algorítmico y brecha de género sea más intensa (empleo, protección social).
- Reforzar la cooperación entre autoridades de IA, de protección de datos y organismos de igualdad, incluida la posibilidad de actuaciones conjuntas en casos de sospecha de sesgos de género.

Sólo una supervisión proactiva, dotada de capacidades técnicas y de sensibilidad igualitaria, puede hacer efectivo el mandato del AI Act de proteger los derechos fundamentales frente a los riesgos de los sistemas de alto riesgo.

10.7. HACIA UN MARCO INTEGRADO DE “DERECHO DEL TRABAJO ALGORÍTMICO” CON PERSPECTIVA DE GÉNERO

En conjunto, la investigación sugiere la necesidad de avanzar hacia un marco integrado de derecho del trabajo algorítmico que incorpore de forma sistemática la perspectiva de género. Esto implica:

- Reinterpretar las categorías clásicas del Derecho del Trabajo –poder de dirección, control empresarial, deber de información, tutela antidiscriminatoria– a la luz de los nuevos dispositivos algorítmicos.
- Ajustar las normas de protección de datos y de IA para que dialoguen con las categorías propias del derecho antidiscriminatorio y de la igualdad de género.
- Reforzar la negociación colectiva y la participación de la representación de las personas trabajadoras como contrapeso democrático frente a la automatización de decisiones laborales.

- Situar la política de igualdad de género en el centro de las estrategias de digitalización del trabajo, tanto en la empresa como en las agendas públicas.

La conclusión de fondo es clara: la transición digital del trabajo y la penetración de la IA en las relaciones laborales no son procesos inevitables que el Derecho deba limitarse a acompañar, sino campos de disputa normativa y política en los que se juega la efectividad del principio de igualdad entre mujeres y hombres. Aceptar la neutralidad de los algoritmos equivale a renunciar a décadas de avances en igualdad laboral; someter la IA a las exigencias del Estado social y democrático de Derecho, con una mirada feminista, es condición indispensable para que la innovación tecnológica sea compatible con la justicia de género.

11. ANEXO BIBLIOGRAFÍA

DOCTRINA

ALLOWAYFI, R.; DENNEHY, D.; DWIVEDI, Y.; COTTERELL, D., “A Systematic Review of Algorithmic Bias in Decision-Making”, Proceedings of SaudiCIS 2024.

BRUSSEVICH, M.; DABLANORRIS, E.; KAMUNGE, C.; KARNANE, P.; KHALID, S.; KOCHHAR, K., Gender, Technology, and the Future of Work, IMF, disponible en: www.imf.org/-/media/files/publications/sdn/2018/sdn1807.pdf. [1]

CASAS BAAMONDE M.E. “Tecnologías, Ciencia, Economía y Derecho”, en Derecho y Tecnologías (Dirs Casas Baamonde y Pérez del Prado), Editorial Universitaria Ramón Areces, 2025.

COBREROS, L.; GALINDO, J.; RAIGADA, T.; ISLA, C., Mujeres en STEM: Desde la educación básica hasta la carrera laboral, Barcelona, EsadeEcPol –Center for Economic Policy, 2024.

COTINO HUESO L.: “Cómo abordar jurídicamente el impacto de la inteligencia artificial en los derechos fundamentales”, en *Derecho y Tecnologías* (Dirs Casas Baamonde y Pérez del Prado), Editorial Universitaria Ramón Areces, 2025.

CRISTOFOLINI CH. “Navigating the impact of AI systems in the workplace: strengths and loopholes of the EU AI Act from a labor perspective”, *Italian Labour Law e-Journal*, Issue 1, Vol. 17 (2024).

GONZÁLEZ GONZÁLEZ, C. S.; GARCÍA, I. y otros, Situación de la investigación en informática en España: Informe Women@Inf, SCIE y Universidad de La Laguna, 2023.

GÓMEZHERRERA, E.; KOESZEGI, S., “Una perspectiva de género sobre la inteligencia artificial y el empleo: el círculo vicioso de la desigualdad digital”, Documento de trabajo 15/2022, Bruegel, disponible en: <https://www.bruegel.org/working-paper/gender-perspective-artificial-intelligence-and-jobs-vicious-cycle-digital-inequality>.

GUDIVADA, V.; APON, A.; DING, J., “Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations”, *International Journal on Advances in Software*, vol. 10, n.º 1&2, 2017.

HOLTZ, H. M.; LEDENDAL, J., “AI Data Governance – Overlaps Between the AI Act and the GDPR”, *Law, Innovation and Technology*, forthcoming 2025.

KAHN, S.; GINTHER, D. K., “Are Recent Cohorts of Women with Engineering Bachelors Less Likely to Stay in Engineering?”, *Frontiers in Psychology*, 6 (2015).

KAMINSKI, M. E., “Binary Governance: Lessons from the GDPR’s Approach to Algorithmic Accountability”, *Southern California Law Review*, 92 (2018): 1529.

MAYORALAS TERRAZAS, I., “Inteligencia artificial y sus sesgos en el análisis de datos: los riesgos de la nueva discriminación”, en *Ensayos Ciberfeministas*, Dykinson, 2023, pp. 227229.

MERCADER UGUINA J.: “Algoritmos e inteligencia artificial en el derecho digital del trabajo” en *Derecho y Tecnologías* (dir. Casas Baamonde y Pérez del Prado), Editorial Universitaria Ramón Areces, 2025.

MERCADER UGUINA J.R. “El derecho de la Unión Europea ante el desarrollo tecnológico: el reglamento de inteligencia artificial”, *Documentación Laboral* núm. 135 (2025).

MERCADER UGUINA J., VELASCO PARDO B.: “El impacto en lo laboral del reglamento de Inteligencia Artificial”, *Revista Derecho Social y Empresa*, núm.. 23 (2025).

MONEREO PÉREZ, J. L., “Vulneración del derecho a la no discriminación y principio de interpretación jurídica conforme a la eficacia de la tutela antidiscriminatoria: salarios de tramitación y despido nulo por embarazo”, *Revista de Jurisprudencia Laboral*, 4 (2024).

RODRÍGUEZ CARDO, I. A., “El Reglamento europeo de inteligencia artificial: una primera valoración sobre su (limitado) impacto laboral”, *Trabajo y Empresa*, n.º 4 (2025).

ROJO TORRECILLA E.: “Nuevas aportaciones sobre el impacto de la inteligencia artificial y la gestión algorítmica, en general, y en el mundo del trabajo en particular” <http://www.eduardorjotorrecilla.es/2026/02/nuevas-aportaciones-sobre-el-impacto-de.html>

ROJO TORRECILLA E.: Sobre tecnología, inteligencia artificial y gestión algorítmica. Impacto en el mundo del trabajo. Presentación, y ordenación y recopilación de artículos publicados en el blog (2020-2026), en http://www.eduardorjotorrecilla.es/2026/02/sobre-tecnologia-inteligencia_11.html

ROMERO RODENAS M.J., *Planes de igualdad, régimen jurídico, negociación y práctica aplicada*, Bomarzo, Albacete, 2026.

SOVRANO, F.; HINE, E.; ANZOLUT, S.; BACCHELLI, A., “Simplifying Software Compliance: AI Technologies in Drafting Technical Documentation for the AI Act”, *Empirical Software Engineering*, 30 (2025), Art. 91.

SURESH, H.; GUTTAG, J. V., “A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle”, *arXiv*, 28 enero 2019.

TRAN, N. et al., “Data Curation and Quality Evaluation for Machine LearningBased Cyber Intrusion Detection”, *IEEE Access*, 10 (2022).

VAN BEKKUM, M.; ZUIDERVEEN BORGESIOUS, F., “Using Sensitive Data to Prevent Discrimination by Artificial Intelligence:

Does the GDPR Need a New Exception?”, *Computer Law & Security Review*, 48 (2023): 105770.

VEALE, M.; VAN KLEEK, M.; BINNS, R., “Fairness and Accountability Design Needs for Algorithmic Support in HighStakes Public Sector DecisionMaking”, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, paper 440 (2018).

ZHA, D. et al., “DataCentric Artificial Intelligence: A Survey”, 2023, disponible en: <https://arxiv.org/pdf/2303.10158>.

“Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning”, *IEEE/ACM Workshop on Gender Equality in Software Engineering*, 2018.

INFORMES, GUÍAS Y DOCUMENTOS INSTITUCIONALES

AEPD, Requisitos para auditorías de tratamientos que incluyan IA, disponible en: <https://www.aepd.es/guias/requisitos-auditorias-tratamientos-incluyan-ia.pdf>.

EDPB, Guidelines 4/2019 on Article 25 – Data Protection by Design and by Default, versión 2.0, 2020, ap. 69.

EUROPEAN COMMISSION, *She Figures 2024: Gender Equality in Research and Innovation*, Brussels, 2024.

EUROPEAN COMMISSION, *She Figures 2024 – Spain Country Fiche*, Brussels, 2024.

EUROSTAT, “Students by field of education (tps00217)”.

INSTITUTO DE LA MUJER, “Alumnado universitario – Mujer-Cifras”, consulta 14 febrero 2026.

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES, *Datos y cifras del Sistema Universitario Español. Curso 20252026*, Madrid, 2025, p. 174.

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES, *Científicas en cifras 2025*, Madrid, marzo 2025, p. 50.

OECD, AI, Data Governance and Privacy. Synergies and Areas of International Cooperation, OECD Artificial Intelligence Papers 22, junio 2024, p. 31.

OECD, “Algorithm and Eve”, disponible en: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/12/algorithm-and-eve_0e889c45/a1603510-en.pdf.

Statista, “Distribución por género de los profesionales de IA y datos en España en 2023”, disponible en: <https://www.statista.com/statistics/1491506/spain-artificial-intelligence-professionals-gender-distribution>.

UGT, Recomendaciones para la negociación colectiva de la IA, Informe del Servicio de Estudios de UGT, noviembre 2023, disponible en: <https://servicioestudiosugt.com/recomendaciones-para-la-negociacion-colectiva-de-la-ia/>.

CCOO, Acuerdo Marco Europeo de los interlocutores sociales sobre digitalización, marzo 2021, disponible en: <https://www.ccoo.es/3ec9e3ddff84034c1a796cb52ac84c09000001.pdf>.

SheAI, “Women in AI: Key Statistics”, disponible en: <https://www.sheai.co/blog/women-in-ai-key-statistics>.

IMK Böckler, documento disponible en: https://www.imk-boeckler.de/data/downloads/OEA/Veranstaltungen/2025/v_2025_10_25_qanas.pdf.

GloCert International, EU AI Act – Technical Documentation (Article 11), guías disponibles en: <https://www.glocertinternational.com/resources/guides/eu-ai-act-technical-documentation-article-11/>.

AI Act Explorer, “Art. 10 Data and data governance”.

Artificial Intelligence Act Explorer, “Artificial Intelligence Act Article 10 (Proposal 25.11.2022)”, disponible en: [https://www.artificial-intelligence-act.com/Artificial_Intelligence_Act_Article_10_\(Proposal_25.11.2022\).html](https://www.artificial-intelligence-act.com/Artificial_Intelligence_Act_Article_10_(Proposal_25.11.2022).html).^[1]

Sitio “Artificialintelligenceact.eu”, Anexo IV, disponible en: <https://artificialintelligenceact.eu/annex/4/>.

AESIA, PresentRIA sistemas de alto riesgo, disponible en: <https://aesia.digital.gob.es/en/presentria-sistemas-de-alto-riesgo>.

NORMATIVA EUROPEA

Carta de Derechos Fundamentales de la Unión Europea.

Directiva 2000/43/CE del Consejo, relativa a la aplicación del principio de igualdad de trato de las personas independientemente de su origen racial o étnico.

Directiva 2000/78/CE del Consejo, por la que se establece un marco general para la igualdad de trato en el empleo y la ocupación.

Directiva 2006/54/CE del Parlamento Europeo y del Consejo, relativa a la aplicación del principio de igualdad de oportunidades e igualdad de trato entre hombres y mujeres en asuntos de empleo y ocupación.

Directiva (UE) 2016/943 del Parlamento Europeo y del Consejo, de 8 de junio de 2016, sobre la protección de los conocimientos técnicos y la información empresarial no divulgados (secretos comerciales).

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016 (Reglamento General de Protección de Datos – RGPD).

Reglamento (UE) 2025/454 de la Comisión, por el que se detalla la composición y funcionamiento del grupo de expertos científicos independientes en IA.

Reglamento Europeo de Inteligencia Artificial (AI Act).

NORMATIVA ESPAÑOLA

Constitución Española de 1978.

Real Decreto Legislativo 2/2015, de 23 de octubre, por el que se aprueba el Texto Refundido de la Ley del Estatuto de los Trabajadores.

Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres.

Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales.

Ley 1/2019, de 20 de febrero, de Secretos Empresariales.

Ley 36/2011, de 10 de octubre, reguladora de la Jurisdicción Social.

Ley 15/2022, de 12 de julio, integral para la igualdad de trato y no discriminación.

Real Decreto Legislativo 5/2000, de 4 de agosto, por el que se aprueba el Texto Refundido de la Ley sobre Infracciones y Sanciones en el Orden Social.

JURISPRUDENCIA Y DOCTRINA JUDICIAL

STC 104/2014, de 23 de junio.

STC 51/2021.

STC 31/2014.

STC 17/2007.

STC 173/2013.

STC 233/2007, de 5 de noviembre, FJ 5.

Declaración del Pleno del Tribunal Constitucional 1/2004, de 13 de diciembre de 2004.

STC de 21 de marzo de 1986.

STS (Sala de lo Social) de 5 de diciembre de 2000 (rec. 4374/1999).

STS de 11 de junio de 2012 (rec. 3336/11).

STS de 13 de julio de 2015 (rec. 221/2014).

STS de 24 de junio de 2024 (rec. 1113/2024).

STS de 21022018 (rec. 842/2016).

STSJ de Madrid (Sala de lo Social, Sección 5.^a), Sentencia 672/2025, de 3 de noviembre de 2025, rec. 350/2025, ECLI: ES:TSJM:2025:13315.

STSJ de Cataluña 6092/2025, de 17 de noviembre de 2025 (rec. 3063/2025).

STSJ de la Comunidad Valenciana, de 8 de octubre de 2025 (rec. 1389/2025).

RECURSOS ELECTRÓNICOS

Agencia Española de Protección de Datos, <https://www.aepd.es> [Consulta: febrero de 2026].

Eurostat – Gender Statistics Database, <https://ec.europa.eu/eurostat/web/gender-statistics> [Consulta: febrero de 2026].

Instituto de las Mujeres – Estadísticas, <https://www.inmujeres.gob.es/MujerCifras/Home.htm> [Consulta: febrero de 2026].

Ministerio de Ciencia, Innovación y Universidades, <https://www.ciencia.gob.es/Estrategias-y-Planes/Estrategias/Mujeres-y-Ciencia.html> [Consulta: febrero de 2026].

OECD Digital Economy Papers, <https://www.oecd.org/digital/digital-economy-papers/> [Consulta: febrero de 2026].

She Figures – European Commission, https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/democracy-and-rights/gender-equality-research-and-innovation/she-figures_en [Consulta: febrero de 2026].



La obra demuestra que la inteligencia artificial aplicada al trabajo no es neutra y puede reproducir o amplificar desigualdades de género preexistentes en el mercado laboral. A partir del marco del AI Act, RGPD y normativa española de igualdad, el estudio analiza cómo la gestión algorítmica de selección, promoción, evaluación y organización del tiempo de trabajo genera riesgos específicos de discriminación directa e indirecta hacia las mujeres. Propone reforzar gobernanza de datos, transparencia y derecho a la explicación, integrar evaluaciones de impacto de género, utilizar la negociación colectiva para cláusulas de no discriminación algorítmica y dotar a las autoridades de control de un mandato explícito de igualdad.



Castilla-La Mancha



Instituto de la Mujer
CASTILLA-LA MANCHA