

EDITORIAL

Refundar lo tuitivo. Garantismo tecnológico y nuevos principios de protección del trabajador en la era digital

Rebuilding the protective approach. Technological guarantees and new principles of worker protection in the digital age

Jesús R. Mercader Uguina

*Catedrático de Derecho del Trabajo y de la Seguridad Social
Universidad Carlos III de Madrid*

ORCID ID: 0000-0001-6301-6788

doi: 10.20318/labos.2026.10639

“El progreso no tiene más adversario que él mismo”

Pascal Chabot, *Téner tiempo. Ensayo de cronosofía*

Resumen: El nuevo garantismo tecnológico laboral constituye la prolongación lógica e histórica del Derecho del Trabajo clásico en un contexto radicalmente transformado por la digitalización y la inteligencia artificial. La fuente del desequilibrio de poder ya no es únicamente la posesión de los medios de producción, sino también el control de los datos, los algoritmos y los sistemas de decisión automatizada. Frente a ello, el texto propone cinco principios fundantes: el principio de prohibición de cosificación algorítmica del trabajador, que erige la dignidad humana como límite absoluto inderogable ante cualquier sistema de IA; el principio de compensación colectiva de la debilidad estructural del consentimiento individual, que convierte la negociación colectiva en el contrapeso institucional frente al poder tecnológico; el principio de proporcionalidad preventiva o desde el diseño de la gestión algorítmica del trabajo, que impone al empleador la carga de justificar ex ante la necesidad e idoneidad del sistema; el principio de transparencia accionable en la gestión algorítmica, que exige una explicabilidad útil para el ejercicio efectivo de derechos y su impugnación; y el principio de reserva de humanidad real en la gestión automatizada del trabajo, que requiere una supervisión humana sustantiva, técnicamente cualificada y con autoridad genuina para oponerse al resultado algorítmico.

Palabras clave: Garantismo tecnológico laboral, inteligencia artificial, gestión algorítmica, cosificación algorítmica, negociación colectiva, proporcionalidad por diseño, transparencia accionable, reserva de humanidad, derechos fundamentales.

Abstract: The new technological labour guarantism constitutes the logical and historical extension of classical Labour Law into a context radically transformed by

digitalisation and artificial intelligence. The source of structural power imbalance is no longer merely the ownership of the means of production, but increasingly the control of data, algorithms, and automated decision-making systems. In response, the text sets out five founding principles: the principle prohibiting the algorithmic reification of the worker, which establishes human dignity as an absolute and non-derogable limit on any AI system; the principle of collective compensation for the structural weakness of individual consent, which transforms collective bargaining into the institutional counterweight to corporate technological power; the principle of preventive or by-design proportionality in algorithmic work management, which places upon the employer the burden of justifying ex ante the necessity and suitability of the system; the principle of actionable transparency in algorithmic management, which demands an explainability that is genuinely useful for the effective exercise of rights and their challenge before the competent authorities; and the principle of real human reserve in automated work management, which requires substantive human oversight that is technically qualified and endowed with genuine authority to oppose and override algorithmic outcomes.

Keywords: Technological labour garantism; artificial intelligence, algorithmic management, algorithmic reification, collective bargaining by-design proportionality, actionable transparency, human reserve, fundamental rights.

I. El derecho del trabajo como proyecto garantista: origen y fundamento

No está de más echar la vista atrás y recordar que el Derecho del Trabajo nació como una rama singular del ordenamiento jurídico fruto de la adaptación de mecanismos jurídicos clásicos, de la redefinición de las construcciones propias del Derecho civil y de la corrección de los efectos más crudos de una relación marcada estructuralmente por la desigualdad entre quien contrata su fuerza de trabajo en situación de necesidad y quien dispone de los medios de producción ⁽¹⁾. Esta rama del ordenamiento ha demostrado, a lo largo de su historia, una notable capacidad de adaptación a los cambios sociales, económicos y tecnológicos; sin embargo, esa misma vocación transformadora no ha ido acompañada de un acuerdo doctrinal o institucional sólido sobre cuál es su razón de ser, cuáles son sus fundamentos, sus fines últimos y el papel que realmente ha desempeñado y debe desempeñar en el ordenamiento jurídico. El Derecho del Trabajo ha sido, en de-

¹ Como con rotundidad señala Otto KAHN-FREUND, *Trabajo y Derecho*, Madrid, MTSS, 1987, p. 52, “el propósito fundamental del Derecho del Trabajo siempre ha sido, y nos atrevemos a decir que siempre lo será, constituir un contrapeso que equilibre la desigualdad de poder negociador que es necesariamente inherente a la relación de trabajo”. Ello configura la verdadera “misión tuitiva” del ordenamiento laboral como la calificaron tempranamente, Gaspar BAYÓN CHACÓN, Eugenio PÉREZ BOTIJA, *Manual de Derecho del Trabajo*, Tomo I, Vol.1, Madrid, Victoriano Suárez, 1957-58, p. 211. Las raíces en la filosofía del Derecho de esta idea pueden encontrarse en Gustav RADBRUCH, que concebía el Derecho social como una disciplina nueva, que opera desde el punto de vista de la protección al débil, como reacción contra el espíritu del Derecho Civil y contra los peligros de la libertad contractual jurídico-formal, concediendo a esta contratación efectos jurídico-públicos. La concepción del Derecho social formulada por el citado autor sirvió para apuntalar la raíz tuitiva del Derecho del Trabajo (Virginia MARTÍNEZ BRETONES, *La filosofía del Derecho de Gustav Radbruch*, Madrid, Fragua, 1994, pp.249-251).

finitiva, más hábil para reaccionar ante las transformaciones externas que para responder a su propia pregunta constitutiva: ¿es un instrumento de protección del trabajador, un mecanismo de gestión del conflicto social, una técnica de organización productiva, o todo ello a la vez?

Esa pregunta, que en otros momentos de la historia pudo responderse con cierta serenidad, adquiere hoy una urgencia inaplazable ⁽²⁾. La digitalización de la economía, la irrupción de la inteligencia artificial (IA), la automatización de procesos y la expansión del trabajo en plataformas no se limitan a exigir ajustes técnicos en este o aquel precepto, sino que vienen a reconfigurar las bases mismas sobre las que se construyó el Derecho del Trabajo. Y es que esta transformación opera en varios planos simultáneos. En el organizativo, la dispersión geográfica de los trabajadores y la ausencia de espacios físicos comunes dificultan la organización sindical y la acción colectiva. En el ideológico, el discurso de la “flexibilidad” y la “autonomía” que acompaña al modelo de plataformas promueve la gestión individual de la relación de trabajo, en la que cada trabajador negocia sus propias condiciones y es percibido como agente independiente. En el estructural, la evidencia empírica confirma una polarización creciente del mercado de trabajo, por un lado, con la expansión de los empleos altamente cualificados vinculados a la tecnología y servicios de baja cualificación difícilmente automatizables y, de otro, con la sombra de la destrucción de los empleos intermedios.

Pero detrás de estas transformaciones, organizativas, ideológicas y estructurales, opera algo más profundo y, en cierto modo, más perturbador como es la insinuación de que el futuro ha de construirse sin memoria ⁽³⁾. La aceleración tecnológica no solo reconfigura las estructuras del empleo; también opera como disolvente sobre la conciencia histórica de la disciplina, presentando la ruptura con el pasado no como una pérdida, sino como una condición necesaria para la adaptación. En ese relato, la continuidad histórica deja de ser un valor para convertirse en una limitación. El Derecho del Trabajo es cuestionado así no solo en sus reglas, sino en su identidad misma. Se le pide que olvide de dónde viene para poder sobrevivir a donde se va.

Para comprender la dimensión real de esta crisis es imprescindible regresar a su fundamento último, porque es precisamente sobre él sobre el que opera con mayor intensidad la presión disolvente de los nuevos relatos tecnológicos. La *ratio essendi* del Derecho del Trabajo, la razón que justificó históricamente su existencia como rama autónoma del ordenamiento, no fue otra que la constatación de la desigualdad estructural inherente a la relación laboral. La simetría contractual predicada por los Códigos Civiles mostraba manifiestas insuficiencias a la hora de proyectar su sombra sobre la realidad social del trabajo por cuenta ajena. Resultaba, por ello, necesario “desmontar la ecuación

² Peter H. DIAMANDIS, Steven KOTLER, *El futuro va más rápido de lo que crees*, Barcelona, Planeta, 2020, p. 29 donde señalan que: “la difusa aceleración que vivimos en la actualidad —o sea, la respuesta al «¿por qué ahora?»— es el resultado de la convergencia de una docena de tecnologías diferentes. Es el progreso a un ritmo que no habíamos visto antes”.

³ Un riesgo del que agudamente advierte Giovanni CAZZETTA, “Una frontera consciente. Derecho del Trabajo y libertad contractual”, en *Estado, juristas y trabajo. Itinerarios del Derecho del Trabajo en el siglo XX*, Madrid, Marcial Pons, 2010, pp. 66-68.

igualitaria de base del Derecho Civil para, ante la desigualitaria e inequitativa realidad social, destacar el carácter necesariamente tutelar de los obreros de las legislaciones de tal nombre después del Derecho del Trabajo y del Derecho Social” (4). Esa constatación fundacional constituyó el punto de partida sobre el que doctrina y jurisprudencia edificaron, durante más de un siglo, un catálogo de principios propios. Calificados con razón como el ADN del sistema, estos principios fueron a un tiempo el fundamento de la autonomía jurídica de la disciplina y la expresión más pura de su función tuitiva y compensadora (5).

Sin embargo, en las últimas décadas ese sistema de garantías clásico ha evidenciado una profunda crisis de eficacia. Los principios aplicativos han ido perdiendo relevancia de forma paralela a la propia maduración del sistema de relaciones laborales: la irrupción de criterios de competitividad empresarial en las normas laborales, la expansión de la negociación colectiva y la disponibilidad normativa progresivamente abierta por el legislador han actuado como factores de sustitución de aquellos principios. Sería, no obstante, un error interpretar esa crisis como el agotamiento de la razón de ser que dio origen a la disciplina. Lo que ha entrado en crisis no es el fundamento, sino su expresión instrumental, esto es, los mecanismos a través de los cuales el ordenamiento laboral articuló históricamente la respuesta al desequilibrio estructural de poder.

La irrupción de la IA, sobre todo a través de la gestión algorítmica, ilustra con particular nitidez esta mutación. Los algoritmos están pasando a convertirse en un instrumento básico en la ordenación y gestión de los poderes empresariales. La IA intensifica el control empresarial y reduce la transparencia de las decisiones, amenazando derechos fundamentales como la privacidad, al tiempo que puede generar ineficiencias y riesgos para la seguridad y la salud si se implanta sin la aportación de quienes conocen el trabajo desde dentro. El problema principal no es tanto la destrucción masiva de empleo, amenaza real, pero cuya magnitud sigue siendo objeto de debate, como el deterioro silencioso de la calidad del trabajo, la automatización de las funciones tradicionales del empleador (contratar, organizar, evaluar, sancionar y despedir) y, en fin, el correlativo incremento del control y la vigilancia en proporciones sin precedentes históricos.

El Derecho del Trabajo no puede afrontar la transformación tecnológica olvidando su historia, sino releýéndola. La reconstrucción que el momento exige no es una ruptura, sino una reinterpretación. Y ello debería llevar a la aplicación de la misma lógica correctora que inspiró los principios clásicos a una realidad en la que la fuente del desequilibrio ya no es únicamente la posesión de los medios de producción, sino también y de manera creciente el control de los datos, los algoritmos y los sistemas de decisión

⁴ María Emilia CASAS BAAMONDE, “La igualdad en las transformaciones y en los orígenes del Derecho del Trabajo”, en G. ROJAS (Coord.), *Orígenes del contrato de trabajo y nacimiento del sistema de protección social*, Albacete, Bomarzo, 2012, p. 86-87.

⁵ De este tema me ocupé ampliamente en Jesús R. MERCADER UGUINA, *Los principios de aplicación del Derecho del Trabajo: Formación, decadencia y crisis* Valencia, Tirant lo Blanch, 2014. Más recientemente, lo ha hecho con gran agudeza, Francisco Javier GOMEZ ABELLEIRA, “Los principios de derecho del trabajo: una aproximación a su fenomenología y a sus funciones”, en AA.VV., *Derecho del Trabajo, cambios económicos y nueva sociedad*, Valencia, Tirant lo Blanch, 2025, pp. 53-82.

automatizada. Lo que ha cambiado es el rostro del poder y lo que no ha cambiado es la necesidad de limitarlo.

II. Fundamentos para un nuevo garantismo tecnológico

La pregunta que surge es inevitable: ¿Qué expresión adoptan los valores constitucionales (la dignidad, la igualdad, la tutela judicial efectiva, la privacidad) cuando el empleador ha delegado en algoritmos e inteligencias artificiales sus funciones más relevantes de gestión, evaluación y dirección?

La respuesta es lo que cabe denominar el nuevo garantismo tecnológico. No se trata de una categoría ajena a la tradición del Derecho del Trabajo, sino de su prolongación lógica e histórica en un contexto radicalmente transformado. Como el garantismo clásico, parte de la constatación del desequilibrio estructural de poder entre las partes de la relación laboral y aspira a compensarlo mediante normas que el trabajador no puede desconocer y que el empleador no puede eludir; pero opera sobre un territorio radicalmente distinto. Este nuevo garantismo exige, ante todo, un cambio de perspectiva interpretativa, leer las nuevas reglas no como expresión de una lógica regulatoria neutral o meramente técnica, sino como manifestación actualizada de la misma voluntad de corrección que dio origen a la disciplina. Los efectos del progreso tecnológico no están predeterminados por la tecnología misma, sino que dependen de cómo se gobierne su despliegue en las organizaciones y de cómo se distribuya el poder en la empresa.

Desde ese cambio de perspectiva, cabe extraer un catálogo de ideas que pudieran servir de base para la construcción de un nuevo garantismo tecnológico frente al desequilibrio de poder generado por la gestión algorítmica del trabajo. Su formulación no pretende ser exhaustiva ni cerrada, el Derecho del Trabajo nunca ha operado sobre catálogos rígidos, sino que se dirige a identificar los ejes sobre los que, entendemos, ha de construirse la protección del trabajador en la era digital.

La primera de esas ideas es que la dignidad del trabajador opera, frente a los sistemas de IA, como límite absoluto y no como variable de ponderación. Los sistemas que reducen al trabajador a un perfil probabilístico opaco, frente al que no cabe réplica ni cuestionamiento, no son simplemente desproporcionados, son ilegítimos con independencia de la eficiencia organizativa que generen. El renovado constitucionalismo digital del trabajo parte de la premisa de que hay un umbral de tratamiento de la persona que no admite justificación funcional alguna, y que ese umbral no puede ser sobrepasado invocando las ventajas del algoritmo. La eficiencia no es una categoría que pueda medirse frente a la dignidad; es una categoría que, necesariamente, debe ceder ante ella. Cabría extraer de ello un principio, el de *prohibición de cosificación algorítmica del trabajador*.

La segunda idea es que la respuesta institucional a la debilidad estructural del consentimiento individual en el contexto laboral no puede ser su sustitución por una pretendida voluntad colectiva, sino en la construcción de fórmulas genuinas de participación organizada. Ante la dependencia económica que convierte la libertad formal del trabajador

en sumisión material, el nuevo garantismo tecnológico sitúa en el centro de su diseño la negociación colectiva como contrapeso institucional al poder tecnológico del empleador, no para legitimar colectivamente lo que individualmente sería inaceptable, sino para que quienes tienen capacidad para representar los intereses de los trabajadores puedan delimitar los ámbitos de uso de la IA, establecer garantías específicas y adaptarlas a cada realidad sectorial o empresarial, sin comprometer irreversiblemente los derechos individuales cuya titularidad permanece en la persona. Ello nos podría llevar a la definición de un *principio de compensación colectiva de la debilidad estructural del consentimiento individual*.

La tercera idea es que el control de la proporcionalidad en el entorno digital no puede seguir operando hacia atrás, sobre el daño ya producido, sino que debe anticiparse al momento del diseño y la implantación. El constitucionalismo digital invierte la lógica clásica del juicio de proporcionalidad y es el empleador quien debe justificar la necesidad antes de actuar, no el trabajador quien debe probar la desproporción después de padecerla. Este desplazamiento de la carga argumentativa, reforzado por las exigencias de limitación de la finalidad y minimización que la normativa de protección de datos ha desarrollado, es la expresión de que el poder tecnológico, como todo poder que incide sobre derechos fundamentales, debe preceder su ejercicio con su justificación. Una nueva realidad que podríamos subsumir en el *principio de proporcionalidad preventiva o desde el diseño de la gestión algorítmica del trabajo*.

La cuarta idea es que la transparencia, en el nuevo garantismo, no es un deber de información sino la condición de posibilidad de todos los demás derechos. Una transparencia que sea genuinamente garantizadora ha de ser explicable en términos accesibles para su destinatario y accionable, es decir, útil para el ejercicio efectivo de derechos, incluida la impugnación. La opacidad de la caja negra no es solo una limitación técnica, es una forma contemporánea de privación de tutela que el constitucionalismo digital no puede tolerar. En definitiva, un *principio de opacidad cero ante el trabajador o, si se prefiere, transparencia accionable en la gestión algorítmica*. Y esa exigencia se prolonga en la quinta idea fundante que no es otra que la supervisión humana de los sistemas de IA ha de ser real y no meramente nominal. El sesgo de automatización (la tendencia a aceptar acríticamente las recomendaciones algorítmicas) es el equivalente tecnológico de la desigualdad estructural clásica. Del mismo modo que el trabajador en situación de necesidad negocia formalmente pero no libremente, el supervisor que carece de capacitación técnica suficiente no puede cuestionar el resultado algorítmico, aunque tenga potestad formal para hacerlo. Ello podría llevar a la definición de un *principio de reserva de humanidad real en la gestión automatizada del trabajo*.

Estas ideas no operan de forma aislada. Son, en suma, los cinco ejes conceptuales de una misma idea, cuya coherencia interna no es casual, sino el resultado de una misma lógica compensadora aplicada al territorio del poder digital. En definitiva, el nuevo garantismo tecnológico no pretende sustituir la memoria del Derecho del Trabajo, sino rescatarla de su disolución. Su apuesta es que la disciplina puede afrontar el futuro sin renunciar a sí misma; no negando la transformación, sino gobernándola desde los valores que le dieron razón de ser. A esa tarea se orienta el análisis que sigue.

III. El principio de prohibición de cosificación algorítmica del trabajador

1. La promesa de eficiencia como nuevo argumento legitimador del poder

En el debate clásico sobre los límites del poder de dirección empresarial, el argumento de la eficiencia organizativa actuaba como contrapeso de los derechos del trabajador, frente al cual los principios del Derecho del Trabajo operaban como límite normativo ⁽⁶⁾. En el contexto contemporáneo, esa función legitimadora ha sido asumida por la eficiencia tecnológica; hoy los sistemas de IA se presentan como instrumentos capaces de valorar a un amplio conjunto de individuos de manera uniforme y, por ello, se reivindican como aliados privilegiados de las políticas de gestión empresarial. La promesa central es la de la hiperlógica algorítmica, sistemas que extraen conclusiones para maximizar la eficiencia y que permitirían, *a priori*, que las decisiones de contratación o de extinción contractual dejen de basarse en corazonadas, impresiones o prejuicios. Esa promesa, con todo, no puede ser ilimitada. La objetividad que el algoritmo reivindica no le inmuniza frente a la discriminación; la impersonalidad que invoca no garantiza el reconocimiento de la individualidad del trabajador; y la eficiencia que promete puede obtenerse, precisamente, a costa de su dignidad.

Dignidad es “una palabra tan excesiva que solo el silencio estaría a su altura”, “una noción sin concepto”, que de obtenerlo incurriría en la “paradoja moral que lleva consigo que su definición lleva consigo también la de la indignidad” ⁽⁷⁾. El Tribunal Constitucional ha rechazado expresamente las concepciones que devalúan a la persona “a mera fuerza de trabajo” por ser incompatibles con el artículo 10.1 CE, imponiendo que la interpretación de las normas laborales se realice a la luz de la dignidad y del libre desarrollo de la personalidad ⁽⁸⁾. La cuestión jurídicamente relevante no es, pues, la eficiencia del sistema —que puede ser real y demostrable—, sino si esa eficiencia se obtiene a costa de quien trabaja. En la era digital, la IA ofrece oportunidades que no cabe ignorar, pero también riesgos que el ordenamiento debe calificar como inaceptables cuando comprometen derechos fundamentales; riesgos cuya materialización depende, en última instancia, del diseño, la orientación y la explotación de los sistemas por actores con poder, esto es, compañías tecnológicas y agentes económicos y políticos ⁽⁹⁾. La denominada neutralidad matemática del algoritmo es, en rigor, una ficción. Todo sistema de IA incorpora las decisiones de quienes lo diseñan y los sesgos de los datos con que se entrena.

⁶ Definía Alfredo MONTOYA MELGAR, en su obra seminal, “*El poder de dirección del empresario*”, dicho poder como “*el conjunto de facultades jurídicas a través de cuyo ejercicio el empresario dispone del trabajo realizado por su cuenta y a su riesgo, ordenando las singulares prestaciones laborales y organizando el trabajo de la empresa*” (Madrid, IEP, 1965, p. 44).

⁷ Expresiones, todas ellas, que pertenecen al brillantísimo estudio de Javier JIMÉNEZ CAMPO, “Artículo 10.1”, en María Emilia CASAS BAAMONDE y Miguel RODRÍGUEZ-PIÑERO BRAVO FERRER (Dir.), *Comentarios a la Constitución. XXX Aniversario*, Madrid, Fundación Wolters Kluwer, 2009, pp.217-218. Si bien es cierto que la utilidad del concepto de dignidad humana ha sido severamente cuestionada. Una crónica de este rechazo puede encontrarse en Ricardo GARCÍA MANRIQUE, “¿Es útil el concepto de dignidad humana?”, *Cronista del Estado Social y Democrático de Derecho*, 2016, n.º 61-62, pp. 14-19.

⁸ STC 192/2003, de 27 de octubre, FJ 7.

⁹ Gregorio CÁMARA VILLAR, *Protección de datos e inteligencia artificial en perspectiva europea*, Manuel ARAGÓN REYES, Juan José SOLOZABAL ECHAVARRIA y Juan Luis REQUEJO PAGÉS (Dir.), *El Estado Constitucional Democrático. Libro en Homenaje a Javier Jiménez Campo*, Madrid, CEPC, 2024, p. 441.

2. La dignidad como límite absoluto en el marco normativo europeo

La dignidad humana no es, ni en el Derecho constitucional español ni en el Derecho de la Unión Europea, una cláusula de carácter meramente retórico. En el plano constitucional, el artículo 10.1 CE proclama la dignidad, junto con los derechos inviolables inherentes a la persona, el libre desarrollo de la personalidad y el respeto a la ley y a los derechos de los demás, como fundamento del orden político y de la paz social. El Tribunal Constitucional la define como un “valor espiritual y moral inherente a la persona, que se manifiesta singularmente en la autodeterminación consciente y responsable de la propia vida y que lleva consigo la pretensión al respeto por parte de los demás” ⁽¹⁰⁾, configurándola además como el “*minimum* invulnerable que todo estatuto jurídico debe asegurar, de modo que, sean unas u otras las limitaciones que se impongan en el disfrute de derechos individuales, no conlleven menosprecio para la estima que, en cuanto ser humano, merece la persona” ⁽¹¹⁾.

La dignidad produce un efecto impeditivo, esto es, prohíbe normas o actos degradantes que traten al ser humano como mero objeto. Decía Kant que, “en el reino de los fines, todo tiene un *precio* o una *dignidad*. Aquello que tiene precio puede ser sustituido por algo *equivalente*; en cambio, lo que se halla por encima de todo precio y, por tanto, no admite nada equivalente, eso tiene dignidad” ⁽¹²⁾. De suerte que el acto inmoral por excelencia será siempre un acto de cosificación, en virtud del cual, confundiendo al hombre con la cosa, lo deshumaniza al dispensar indebidamente a los seres con dignidad el tratamiento de los que solo tienen precio ⁽¹³⁾.

Cuando un sistema de IA genera un *score* de rendimiento, un perfil de riesgo de absentismo o un índice predictivo de rotación laboral, no efectúa únicamente una operación técnica de procesamiento de información; sustituye al trabajador como sujeto por una representación estadística de sí mismo, producida sin su participación, opaca en su elaboración e incontestable en su contenido. Esta sustitución tiene consecuencias jurídicas cuando opera sin posibilidad real de comprensión ni de contradicción por parte del afectado, produce exactamente la objetivación que el artículo 10.1 CE y la jurisprudencia constitucional prohíben. En términos teóricos, la dignidad de la persona no se agota en un supuesto de hecho cerrado, se activa siempre que exista riesgo para la persona y tiende a presentarse como límite no derrotable ⁽¹⁴⁾ o, cuanto menos, como barrera que no debería ceder por simple ponderación utilitarista ⁽¹⁵⁾.

¹⁰ STC 53/1985, de 11 de abril, FJ 8. Para un análisis de las proyecciones de la idea de dignidad en la jurisprudencia del Tribunal Europeo de Derechos Humanos, vid. Consuelo CHACARTEGUI JÁVEGA, *Dignidad de los trabajadores y derechos humanos del trabajo según la jurisprudencia del Tribunal Europeo de Derechos Humanos*, Albacete, Bomarzo, 2013.

¹¹ STC 120/1990, de 27 de junio, FJ 4.

¹² Immanuel KANT, *Fundamentación metafísica de las costumbres*, Madrid, Espasa Calpe, 1983, p. 92, Traducción Manuel GARCIA MORENTE. Y, añade, “pero la legislación misma, que determina todo valor, debe por eso justamente tener una dignidad, es decir, un valor incondicionado, incomparable, para el cual solo la palabra *respeto* ofrece la expresión conveniente de la estimación que un ser racional debe tributarle” (también, *Fundamentación...*, p. 94).

¹³ Javier GOMÁ LANZÓN, *Dignidad*, Madrid, Galaxia Gutenberg, 2019, p. 28

¹⁴ Cuando una obligación no cede ante ninguna circunstancia, se afirma que tiene carácter “absoluto o inderrotable”, Robert ALEXI, *Teoría de los derechos fundamentales*, Madrid, CEC, 2007, 2ª ed., p.262.

¹⁵ Manuel ATIENZA, *Sobre la dignidad humana*, Madrid, Trotta, 2020, p. 73.

La referencia kantiana que subyace a este razonamiento no es decorativa. La segunda formulación del imperativo categórico –“obra de tal modo que uses la humanidad, tanto en tu persona como en la persona de cualquier otro, siempre como un fin al mismo tiempo y nunca solamente como un medio”⁽¹⁶⁾– articula una prohibición que la dogmática constitucional ha asumido plenamente. Sin embargo, y como advierte Atienza, esa fórmula no es exclusivamente negativa ni puramente formal. Y es que lo vedado no es cualquier tratamiento instrumental de la persona, inevitable en alguna medida en toda relación laboral, sino el tratamiento *exclusivamente* instrumental, aquel que no reconoce al sujeto ninguna dimensión más allá de su utilidad para el sistema. El algoritmo no conoce al trabajador; conoce sus datos. Y en esa sustitución de la persona por sus datos opera la cosificación constitucionalmente proscrita.

En el Derecho de la Unión Europea, el Considerando 48 del Reglamento (UE) 2024/1689, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de IA (en adelante, RIA) precisa que la magnitud de las consecuencias adversas de un sistema de IA para los derechos fundamentales protegidos por la Carta –entre los que enumera expresamente la dignidad humana, el respeto de la vida privada y familiar, la protección de datos, la libertad de expresión e información, la libertad de reunión y asociación, la no discriminación y los derechos de los trabajadores– es determinante a efectos de la clasificación de dicho sistema como de alto riesgo. Esta jerarquía no es meramente programática, antes al contrario, activa un conjunto de obligaciones concretas –transparencia, supervisión humana, evaluación de impacto y rendición de cuentas–, así como deberes de pruebas y controles *ex ante* y *ex post*, documentación y gestión de riesgos durante todo el ciclo de vida del sistema. En el plano del Derecho internacional convencional, el Convenio Marco del Consejo de Europa sobre IA y Derechos Humanos, Democracia y Estado de Derecho (publicado en el Diario Oficial de la Unión Europea el 13 de mayo de 2026 y primer instrumento internacional jurídicamente vinculante en la materia) establece, en su artículo 1, el objetivo de garantizar la plena coherencia del ciclo de vida de los sistemas de IA con los derechos humanos, la democracia y el Estado de Derecho, e impone, en su artículo 7, el deber de cada Parte de adoptar o mantener medidas que respeten la dignidad humana y la autonomía individual en relación con dichas actividades.

La articulación entre la CE, la Carta de Derechos Fundamentales de la UE, el RIA y el Convenio del Consejo de Europa forma, así, un marco normativo de protección reforzada que sitúa a la dignidad como límite transversal e ineludible de toda gestión algorítmica del trabajo.

¹⁶ Cualquier acción o práctica, para ser considerada moral para Kant, debe cumplir los tres siguientes requisitos (“imperativo categórico”). En primer lugar, una acción será moralmente permisible sólo si cualquier persona puede cumplirla en cualquier momento y sigue teniendo sentido. La segunda prueba exige que se trate a los seres humanos como fines valiosos en sí mismos y no meramente como instrumentos para conseguir un fin y como dignos de respeto y dignidad. El tercer test de Kant requiere que se respete la autonomía de los seres humanos. De ahí que, para analizar si una acción es moral, habría que pedir a un tercero que la juzgase si la persona estuviera dispuesta a ser su agente, pero no su receptor, la conducta es inmoral. Immanuel KANT, *Fundamentación de la metafísica de las costumbres*. cit.

3. *El cuerpo del trabajador como nuevo espacio de control*

De entre los bienes jurídicos que ese marco de protección reforzada ampara, ninguno resulta tan directamente interpelado por la expansión de la gestión algorítmica como la corporalidad misma del trabajador. En este espacio la dignidad no es ya una abstracción normativa, sino una realidad física, inalterable e irrenunciable. Es precisamente por ello que en este ámbito la idea de dignidad adquiere su dimensión más exigente cuando la tecnología no se dirige a los actos del trabajador, sino a su propio cuerpo. La proliferación de técnicas de biometría laboral –huellas dactilares, reconocimiento facial y vocal, geolocalización integral, implantes y neurotecnologías– ha desplazado el eje del control empresarial hacia la corporalidad misma de la persona trabajadora ⁽¹⁷⁾. Este salto tecnológico es cualitativamente distinto de formas anteriores de supervisión, puesto que la fuente corporal es, por naturaleza, permanente e insustituible, aunque los datos se borren, los rasgos físicos que los generan permanecen. Ello permite la construcción de perfiles de salud, rasgos psíquicos o étnicos y estados emocionales, anticipando escenarios de vigilancia ubicua que penetran en lo que he denominado la intimidad de la intimidad.

Cuando la gestión algorítmica penetra en ese ámbito corporal –midiendo el pulso cardíaco, analizando microexpresiones faciales o rastreando en tiempo real cada movimiento del trabajador– traspasa el umbral de lo que el ordenamiento protege como intimidad personal. El riesgo cualitativo que genera la acumulación de datos biométricos supera con creces el derivado de cualquier otro tipo de datos. El trabajador no puede sustraerse a su propia corporalidad, no puede cambiar su iris, su voz o su marcha para protegerse de un sistema que los procesa de forma permanente. A ello se añade la dimensión psicológica de la vigilancia continua ⁽¹⁸⁾, lo que se materializa en la capacidad de los dispositivos inteligentes para analizar constantemente la conducta del trabajador lo que lleva consigo inducir comportamientos de adaptación y dependencia tecnológica que afectan tanto a su desempeño profesional como a su bienestar personal. Esta dinámica reproduce los efectos que Michel Foucault describió a partir de la metáfora del panóptico de Bentham, de modo que la mera posibilidad de ser observado en todo momento basta para modificar el comportamiento del observado, con independencia de que la observación se produzca o no efectivamente. El Preámbulo del Convenio del Consejo de Europa se muestra expresamente contrario al uso de sistemas de IA “con fines represivos”, rechazando en particular las “prácticas arbitrarias o ilegales de vigilancia y censura que minan la privacidad y la autonomía individual”.

¹⁷ Sobre este tema me remito a mi estudio: “El “Big Bang” de la biometría laboral. De la huella dactilar a los neurodatos”. *Labos: Revista de Derecho del Trabajo y Protección Social*, 2024, Volumen 5, Número nº 2, pp.4-23.

¹⁸ Como señala Carlos CASTILLA DEL PINO, *Arquitectura de la vida humana*, Madrid, Real Academia Española de la Lengua, 2006, p. 29, “las actuaciones íntimas son estrictamente mentales. A la vida íntima pertenecen pensar en, fantasear con, imaginar que, dar paso a nuestras emociones y sentimientos, a nuestros recuerdos, deseos, intenciones, que no podemos hacer públicos y si se pudieran aparecerían desmesurados e intolerables”. El cuerpo y la conciencia se encuentran implicados en vivencias y sentimientos compartidos, de modo que el sistema nervioso, el cerebro y el modo de comportarse y afrontar el mundo no pueden ser comprendidos como estructuras separadas. Antes, al contrario, como dijo MERLEAU-PONTY en su *Fenomenología de la percepción*, deben ser entendidos como una estructura sistémica.

Convertir el cuerpo del trabajador, con sus patrones biométricos y sus estados emocionales, en objeto de extracción de datos al servicio de la productividad empresarial es, precisamente, la forma de cosificación que la CE prohíbe y que la STC 181/2004, de 2 de diciembre, FJ 13 define como el trato que “rebaja al individuo a un nivel material o animal” o que lo “mediatiza o instrumentaliza, olvidándose de que toda persona es un fin en sí mismo”. En coherencia con este marco, el artículo 11 del Convenio del Consejo de Europa sobre IA impone a las Partes el deber de adoptar medidas para garantizar que las actividades en el ciclo de vida de los sistemas de IA protejan los derechos a la intimidad y los datos personales de las personas afectadas, con salvaguardias efectivas a su favor.

4. La dignidad como norma de validez absoluta: del límite interpretativo a las prohibiciones irrenunciables

La dignidad de la persona no opera únicamente como límite interpretativo ponderable con otros bienes o principios en cada caso concreto. La transición digital en el lugar de trabajo no es un proceso neutro es, realmente, una transformación del poder –de quién lo ejerce, cómo y con qué medios– que exige ser orientada por principios que protejan a quienes son objeto de ese poder. En determinados supuestos, la dignidad de la persona cristaliza en prohibiciones absolutas, esto es, en exigencias que no admiten legitimación bajo ninguna justificación de eficiencia, del mismo modo que ciertos derechos laborales no pueden ser objeto de renuncia ni siquiera consentida. La diferencia entre uno y otro plano no es de grado, sino de naturaleza. Frente a un límite ponderable, el argumento contrario puede prevalecer si resulta suficientemente fuerte; frente a una prohibición absoluta, el argumento de eficiencia no tiene acceso a la ponderación porque el ordenamiento ha decidido de antemano que determinados usos de la tecnología sobre la persona son, en todo caso, incompatibles con su dignidad.

En el contexto de la gestión algorítmica del trabajo, la dignidad de la persona opera, además, no solo como criterio hermenéutico de otras normas, sino como norma de validez material de todo sistema de IA empleado en el ámbito laboral. Concebidos de un modo sustantivo, los principios constitucionales del artículo 10.1 CE condicionan el contenido mismo de las reglas y los actos jurídicos por ellos afectados, con consecuencias que es posible enunciar con precisión. Todo sistema de IA que genere decisiones sobre trabajadores sin posibilidad real de comprensión ni de contradicción vulnera la dignidad de la persona, por operar la cosificación que la STC 181/2004 prohíbe. Toda decisión empresarial apoyada en sistemas que reduzcan al trabajador a un perfil probabilístico incontestable viola el mandato de respeto a su condición de ser racional, igual y libre que el artículo 10.1 CE impone. Y el argumento de eficiencia no puede, por sí solo, justificar ninguna de esas vulneraciones, pues allí donde la dignidad opera como límite absoluto, no cede frente a ninguna ponderación de utilidad. Esta dimensión normativa de la dignidad –como condición de validez y no meramente como parámetro interpretativo– es la que otorga fundamento constitucional a los instrumentos regulatorios que, en el plano

del Derecho de la Unión y del Derecho internacional, han trasladado esa exigencia a un catálogo de prohibiciones positivas.

Esta lógica –la de los límites cualitativos frente a la eficiencia– es la que inspira el catálogo de prácticas prohibidas que el artículo 5 del RIA establece con carácter absoluto. El Reglamento prohíbe expresamente la introducción en el mercado, la puesta en servicio o la utilización de sistemas de IA destinados a inferir las emociones de una persona física en el lugar de trabajo, así como los sistemas de categorización biométrica que clasifiquen individualmente a las personas sobre la base de sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida u orientación sexuales. Se prohíben igualmente los sistemas de predicción o evaluación de riesgo basados en perfiles o rasgos de personalidad, por vulnerar la dignidad humana y la presunción de inocencia. El artículo 16.4 del Convenio del Consejo de Europa sobre IA encuadra estas prohibiciones en una lógica más amplia, según la misma cada Parte evaluará la necesidad de moratorias, prohibiciones u otras medidas adecuadas respecto de usos de sistemas de IA que resulten incompatibles con el respeto de los derechos humanos, la democracia o el Estado de Derecho. Las prohibiciones absolutas del artículo 5 del RIA son, precisamente, la traducción positiva de ese mandato. El legislador europeo ha calificado determinados usos de la IA en el trabajo como incompatibles con la dignidad humana, sustrayéndolos por completo al ámbito de la ponderación caso a caso.

En esta misma dirección, el Parlamento Europeo ha articulado un catálogo complementario de prácticas igualmente vedadas, con vocación de completar el perímetro protector trazado por el RIA. La Resolución de 17 de diciembre de 2025, sobre digitalización, inteligencia artificial y gestión algorítmica en el lugar de trabajo, propone, en su Recomendación 5, que se prohíba la recogida, el almacenamiento, el seguimiento, la venta o cualquier otro tratamiento de un conjunto de categorías de datos cuya explotación resulta, por su propia naturaleza, incompatible con el estatuto de la persona trabajadora. El elenco abarca los datos relativos a los estados emocionales, psicológicos o neurológicos de los trabajadores, a su actividad cognitiva o a sus parámetros biométricos; las comunicaciones privadas, incluidas las mantenidas con otros trabajadores y con los representantes de los trabajadores, salvo que su tratamiento resulte lícito al amparo del artículo 6, apartado 1, letras c) o d), del RGPD; los datos generados cuando el trabajador se encuentra fuera de servicio, no ofrece ni realiza trabajo o se halla en espacios privados, así como los datos de geolocalización en tiempo real obtenidos fuera del horario laboral o más allá de lo estrictamente necesario para la ejecución de las tareas; los datos destinados a predecir el ejercicio de derechos fundamentales, en particular la libertad de asociación, el derecho de negociación y acción colectivas o el derecho a la información y consulta; y, en fin, las categorías especiales de datos personales contempladas en el artículo 9, apartado 1, del RGPD, así como cualquier inferencia sobre dichas categorías. La Resolución precisa, además, que estas prohibiciones han de proyectarse sobre todo el procedimiento de contratación, extendiendo así la tutela al momento anterior al inicio de la relación laboral.

La lectura conjunta de ambos textos –el artículo 5 del RIA y la Recomendación 5 de la Resolución– permite trazar con mayor nitidez el contorno de lo absolutamente prohibido. No se trata de una lista cerrada de tecnologías concretas, sino de la delimitación de un espacio de inmunidad del trabajador frente a determinadas formas de escrutinio digital que, con independencia de su grado de sofisticación técnica o de la finalidad organizativa alegada, resultan radicalmente incompatibles con la dignidad de quien trabaja. La prohibición no admite modulación por razón de utilidad, ni siquiera cuando el tratamiento de datos se presente revestido de una lógica de mejora de la productividad o de gestión de riesgos empresariales. El ordenamiento ha sustraído ese ámbito al cálculo de eficiencia y lo ha trasladado al terreno de las exigencias deontológicas absolutas. Ello no es sino la consecuencia lógica de reconocer a la dignidad su dimensión de norma de validez. Si un sistema de IA no supera el test de compatibilidad con el artículo 10.1 CE, ninguna consideración de eficiencia puede suplir esa carencia, porque la validez no es un atributo que se conquiste por acumulación de utilidades, sino una condición previa que el ordenamiento impone con carácter incondicionado.

IV. El principio de compensación colectiva de la debilidad estructural del consentimiento individual

1. El consentimiento como categoría poliédrica

El consentimiento ha ocupado siempre un lugar central en la definición del contrato de trabajo hasta el punto de haber sido configurado como un contrato en el que es necesario el “consentimiento continuo” del trabajador ⁽¹⁹⁾. En su formulación tradicional, el consentimiento representa la manifestación más genuina de la autonomía de la voluntad. Quien consiente no se limita a asentir, sino que decide y se afirma a sí mismo frente al otro en plenitud de su condición de sujeto libre. Se trata de una noción que hunde sus raíces en el pensamiento contractualista liberal y que ocupa un lugar central en el proyecto político moderno, edificado sobre la convicción que el derecho hace suya de que los individuos adultos somos capaces de obligarnos libremente ante los demás y ante el poder público. En ese marco, el consentimiento cumple una función política de primer orden, pues permite trazar la línea que separa el ejercicio legítimo de la autoridad de la imposición arbitraria, distinguiendo el orden social nacido del acuerdo libre entre personas iguales del dominio que se sostiene únicamente sobre la fuerza y la coacción. De él depende nada más y nada menos que la diferencia entre la libertad y la sumisión ⁽²⁰⁾.

Sin embargo, el consentimiento encierra, como el dios Jano, una doble faz. Su otra cara se proyecta sobre relaciones de desigualdad y estructuras que dominan a los sujetos. La idea del consentimiento, por tanto, caracteriza dos situaciones muy diferentes:

¹⁹ Manuel ALONSO OLEA, *Introducción al Derecho del Trabajo*, Madrid, EDERSA, 1962, pp. 136-137.

²⁰ Clara SERRA, *El sentido de consentir*, Madrid, Anagrama, 2024, pp. 13-14.

“[...] aquella de la relación de fuerza y de su salida impuesta (ceder antes que consentir), y aquella del contrato entre partes más o menos iguales (o consentimiento mutuo)”⁽²¹⁾. Lo relevante en los casos que nos ocupan es que la posibilidad de debilidad de uno de los contratantes afecta a su libertad contractual, de modo que puede consentir un contrato sustancialmente injusto, o bien forzado por las circunstancias, o bien sin ser enteramente consciente⁽²²⁾. Por ello, cuando ese consentimiento se traslada al ámbito de las relaciones laborales y se enfrenta a los desafíos propios de la transformación digital, su virtualidad garantizadora se revela profundamente comprometida. La proliferación de sistemas de IA, el tratamiento masivo de datos personales y la digitalización de los entornos de trabajo han convertido al consentimiento en una categoría sometida a una tensión estructural que exige ser analizada en sus múltiples dimensiones.

Su análisis exige partir de una premisa ineludible y que no es otra que el hecho de que el *minimum* invulnerable de la dignidad no es negociable⁽²³⁾. No puede ser objeto de renuncia y no puede ser sustituido por la adhesión del trabajador a los términos de uso de ningún sistema tecnológico, por sofisticado que este sea. Esta premisa es el fundamento normativo sobre el que descansa todo el edificio del Derecho del trabajo. A partir de ella, es posible desplegar el análisis de las diferentes dimensiones que el consentimiento adquiere en el contexto de la transformación digital.

2. La paradoja estructural del consentimiento en la relación de subordinación

El principio de indisponibilidad de derechos laborales nació como respuesta a una necesidad estructural a la necesidad de compensar el desequilibrio de poder que define, en su esencia, la relación entre empleador y trabajador⁽²⁴⁾. Su lógica era tutelar al más débil impidiéndole renunciar a aquellas protecciones que el ordenamiento jurídico, mediante norma imperativa, le reconocía como inderogables. La transformación digital introduce, sin embargo, una inversión paradójica en esta lógica tradicional pues ya no es el trabajador quien puede verse tentado a renunciar a sus derechos, sino que es el empleador quien requiere su consentimiento para legitimar tratamientos de datos de alto impacto, presentando dicho consentimiento como expresión auténtica de la autonomía individual del trabajador. La asimetría estructural de la relación laboral permanece intacta, pero su manifestación formal ha cambiado de dirección. El instrumento concebido históricamente para proteger al trabajador de sí mismo, es decir, de sus propias renunciadas inducidas por

²¹ Geneviève FRAISSE, *Del consentimiento*, Palinodia, Santiago de Chile, 2007. Traducción de Felipe Alarcón

²² Para una reflexión que excede el estricto plano laboral, vid. Esther GOMEZ CALLE, *Desequilibrio contractual y tutela del contratante débil*, Pamplona, Aranzadi, 2018, p. 32.

²³ Como muestra de ello, la Recomendación sobre la Ética de la Neurotecnología de la UNESCO de 11 de noviembre de 2025, en su apartado (134) sostiene que "en principio, el consentimiento, por sí mismo, no debería ser suficiente como sola base jurídica de cualquier procesamiento intrusivo con empleo de neurotecnología".

²⁴ Nos permitimos volver a remitir a Jesús R. MERCADER UGUINA, *Los principios de aplicación del Derecho del Trabajo: Formación, decadencia y crisis*, cit. y, muy principalmente, a las consideraciones de Aurelio DESDENTADO BONETE, en el *Prólogo*.

la presión contractual, se transforma ahora en el mecanismo a través del cual esa misma presión encuentra una nueva vía de legitimación.

Esta paradoja adquiere una dimensión constitucional que no puede soslayarse. El artículo 10 de la Constitución Española (en adelante, CE) ancla el reconocimiento de la dignidad humana en la capacidad de la persona para autorregular sus propios marcos de intereses. Como ha señalado Díez-Picazo, el reconocimiento de un poder de autonomía privada o de creación normativa privada está directamente enraizado en dicho precepto constitucional, pues “solo se reconoce la dignidad de la persona si se le permite autorregular sus marcos de intereses” (25). Desde esta perspectiva, la indisponibilidad de derechos se convierte en una categoría de justificación compleja cuando su aplicación implica proteger al trabajador incluso frente a su propia voluntad declarada, pues ello supondría, en cierta medida, negar la capacidad de autodeterminación que la misma CE reconoce como fundamento de la dignidad personal.

No obstante, la respuesta a esta tensión no puede consistir en una simple ampliación del espacio disponible a la autonomía contractual, concebida esta como panacea frente a la rigidez normativa. Como advirtiera Rodríguez-Piñero, “la revalorización del contrato debe servir para la tutela de la persona y la dignidad del trabajador y de sus derechos fundamentales, evitando que instrumentos como la indisponibilidad de derechos, que inicialmente tenían la finalidad de asegurar la autodeterminación del trabajador, degeneren en regulaciones cerradas que suplanten y supriman la voluntad contractual” (26). La atenuación de la normativa heterónoma y la apertura de mayores espacios para la disponibilidad de derechos han de orientarse, en consecuencia, a la efectiva realización de la dignidad del trabajador, y no a su erosión encubierta bajo el ropaje formal del consentimiento.

El entorno digital agudiza esta problemática de manera singular (27). Las decisiones sobre el tratamiento de datos personales en el ámbito laboral involucran con frecuencia dimensiones que afectan directamente a la esfera más íntima del trabajador. Cuando el consentimiento se presta en el marco de una relación de dependencia económica y jerárquica, bajo la presión difusa pero constante del vínculo laboral, su capacidad para expresar una autodeterminación genuina queda seriamente comprometida.

²⁵ Luís Díez-PICAZO, “La Ley y la potestad normativa privada”, en Antonio LÓPEZ PINA, *Democracia representativa y parlamentarismo*, Madrid, 1994, p. 83. La Ley Fundamental de Bonn tampoco contiene una garantía expresa de la libertad contractual. Según la jurisprudencia del Tribunal Constitucional, su ejercicio dentro de los límites del “orden constitucional” está en todo caso tutelado por el artículo 2.1 de la Ley Fundamental, como proyección del libre desarrollo de la personalidad, en la medida que no lo garanticen otras disposiciones de los derechos fundamentales. Como pone de manifiesto, Konrad HESSE, *Derecho Constitucional y Derecho Privado*, Madrid, Civitas, 1995, p. 55.

²⁶ Miguel RODRÍGUEZ-PIÑERO, “El principio de indisponibilidad de los derechos laborales y el nuevo papel de la autonomía contractual”, en Luis Enrique de la VILLA GIL, Lourdes LOPEZ CUMBRE (Dir.), *Los principios del Derecho del Trabajo*, Madrid, CEF, 2003, p. 126.

²⁷ Henar ÁLVAREZ CUESTA, “El consentimiento individual y su alcance en la inteligencia artificial aplicada al ámbito laboral”, *Documentación Laboral*, 2022, nº 126, pp. 51-68. También, Eduardo GONZÁLEZ BIEDMA, “Derecho a la información y consentimiento del trabajador en materia de protección de datos”. *Temas laborales*, 2017, nº 138, pp. 223-247. Extensamente para el resto de los ámbitos, David AVIÑÓ BELENGUER, *El consentimiento para el tratamiento de datos personales. Nuevos desafíos en la era del Big Data y de la Inteligencia Artificial*, Valencia, Tirant Lo Blanch, 2026,

Esta misma lógica ha encontrado un reflejo preciso en el ámbito de la protección de datos personales, donde el consentimiento del trabajador posee, con carácter general, un valor residual. El propio Reglamento UE/2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de sus datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (corrección de errores DOUE L 127 de 23 de mayo de 2018) (en adelante, RGPD); reconoce, en su Considerando 43, que existe un desequilibrio claro entre el interesado y el responsable del tratamiento que sitúa al trabajador en una posición de vulnerabilidad estructural. Sobre esta base, el Grupo de Trabajo del artículo 29 –actualmente integrado en el Comité Europeo de Protección de Datos– concluyó que “se produce un desequilibrio de poder” en la relación laboral, de tal suerte que “dado que el empleado depende del empleador, no es probable que el interesado pueda negar a su empleador el consentimiento para el tratamiento de datos sin experimentar temor o riesgo real de que su negativa produzca efectos perjudiciales”, siendo por ello “poco probable que un empleado pudiera responder libremente a una solicitud de consentimiento de su empleador”⁽²⁸⁾

Una de las cuestiones más debatidas durante la tramitación del Proyecto de Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales fue la del papel que debía corresponder al consentimiento del trabajador como título habilitante del tratamiento de sus datos en el marco de la relación laboral ⁽²⁹⁾. Aunque la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (en adelante, LOPD).no cambió los parámetros fundamentales de la situación,

²⁸ Grupo de trabajo del artículo 29, Directrices sobre el consentimiento en el sentido del Reglamento (UE) 2016/679, adoptadas el 28 de noviembre de 2017, revisadas por última vez y adoptadas el 10 de abril de 2018, WP 259 rev.01, p. 7. Especialmente significativo es, en este punto, el Considerando (39) de la Directiva (UE) 2024/2831 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, relativa a la mejora de las condiciones laborales en el trabajo en plataformas, cuando señala: “Las personas que realizan trabajo en plataformas a menudo no tienen una verdadera libertad de elección o no pueden denegar o retirar el consentimiento sin perjudicar su relación contractual, dado el desequilibrio de poder existente entre la persona que realiza trabajo en plataformas y la plataforma digital de trabajo. Por consiguiente, las plataformas digitales de trabajo no deben tratar los datos personales de una persona que realiza trabajo en plataformas en razón de que esta haya dado su consentimiento para el tratamiento de sus datos personales”.

²⁹ Diversas enmiendas presentadas por el Grupo Parlamentario Confederal de Unidos Podemos-En Comú Podem-En Marea al Proyecto de Ley Orgánica de Protección de Datos de Carácter Personal pusieron de manifiesto una posición clara y coherente: en el ámbito de las relaciones laborales, el consentimiento del trabajador no puede operar como título legítimo de habilitación del tratamiento de datos. Sobre esta base, la Enmienda núm. 7 propuso que los datos relativos a la salud de los trabajadores no pudieran ser objeto de acceso ni comunicación a terceros, estableciendo expresamente que el mero consentimiento del interesado no bastaría para levantar esta prohibición, mientras que la Enmienda núm. 9 planteó la supresión íntegra del artículo relativo al tratamiento de datos de contacto de los trabajadores, al entender que dicho precepto permitía que la empresa cediera a terceros los datos de localización profesional del trabajador sin necesidad de que este otorgara su consentimiento, y que el tratamiento de esos datos debía contar con el consentimiento del interesado, tanto y más cuando se pretendiera su comunicación a terceros. En materia de videovigilancia, la misma línea restrictiva se plasmó en la Enmienda núm. 13, que propuso una redacción del artículo 22.5 en la que se establecía con rotundidad que el consentimiento otorgado por los trabajadores o por sus representantes no bastaría en ningún caso para alterar lo establecido en el precepto, al tiempo que se prohibía expresamente la captación de imágenes para el control directo o indiscriminado de los trabajadores. Sobre la referida tramitación, Francisco MARTÍNEZ VÁZQUEZ, *La tramitación parlamentaria de la Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales*. RALLO LOMBARTE, Artemi (director). Tratado de datos personales. Tirant lo Blanch, 2019, pp.53-77.

lo cierto es que avivó el debate hasta el punto de que la Agencia Española de Protección de Datos (en adelante, AEPD) llegó a afirmar en su Resolución R/0041/2019 que “los trabajadores no están nunca en condiciones de dar, denegar o revocar el consentimiento libremente, habida cuenta de la dependencia que resulta de la relación” ⁽³⁰⁾. Desde el ámbito del derecho de la protección de datos se confirma así, con toda claridad, que el consentimiento formalmente prestado por el trabajador no puede sustituir al análisis sustantivo de las condiciones reales en que dicho consentimiento se ha formado y manifestado.

La clave del problema reside, por tanto, en la distinción rigurosa entre el consentimiento que expresa una auténtica autodeterminación del trabajador, aquel que merece ser respetado y protegido como manifestación de su dignidad, y el consentimiento que, bajo esa misma apariencia formal, constituye una nueva forma de sometimiento al poder del empleador. Esta distinción no es meramente técnica ni puede resolverse a través de criterios puramente formales de validez negocial. Exige, por el contrario, una aproximación sustantiva que atienda a las condiciones materiales en que el consentimiento se produce, a la existencia de alternativas reales para el trabajador, a la proporcionalidad de la finalidad perseguida por el empleador y al grado en que la decisión adoptada afecta a derechos fundamentales irrenunciables.

En definitiva, la autonomía de la voluntad en la relación laboral digital no puede ser ni absolutamente negada, so pena de incurrir en un paternalismo constitucionalmente cuestionable, ni plenamente equiparada a la autonomía que opera entre sujetos en posición de igualdad. La función del ordenamiento jurídico consiste, precisamente, en garantizar que el espacio abierto a la voluntad del trabajador sea un espacio de libertad real y no de sumisión revestida de consentimiento.

3. El convenio colectivo como respuesta institucional a la debilidad individual: El ejemplo de la protección de datos

Ante la inoperancia del consentimiento individual, el ordenamiento jurídico articula una respuesta institucional que no consiste en prescindir de cualquier forma de participación de los trabajadores en la definición de las condiciones de tratamiento, sino en trasladar esa participación al plano colectivo. Cuando los representantes de los trabajadores negocian con el empleador las condiciones de un tratamiento de datos, no están prestando un consentimiento en nombre de cada uno de los trabajadores afectados. El

³⁰ La posición de la AEPD, sin embargo, no está exenta de crítica por su radicalidad. Ha de tenerse en cuenta que el consentimiento explícito del trabajador constituye, en el esquema del RGPD, uno de los títulos habilitantes para levantar la prohibición de tratamiento de las categorías especiales de datos (artículo 9.2.a) del RGPD), y que el legislador español, como acabamos de señalar durante la tramitación parlamentaria de la LOPD, optó expresamente por eliminar diversas limitaciones al uso del consentimiento que habían sido propuestas. Esta tensión no puede resolverse mediante una negación absoluta del valor del consentimiento en el ámbito laboral, pues, como recordara DÍEZ-PICAZO, “la libertad contractual, lejos de ser pura y virginal, es voluntad razonable dentro de un marco razonable de libertad”. La cuestión no es, pues, si el consentimiento puede tener algún valor en la relación laboral, sino en qué condiciones materiales ese valor puede reconocérsele.

convenio colectivo no es el consentimiento individual multiplicado por el número de trabajadores, ni tampoco una delegación de la voluntad individual en un tercero, sino algo cualitativamente diferente: la expresión organizada de los intereses colectivos de un grupo, articulada por quienes tienen capacidad, información y legitimidad para representar esos intereses en un proceso de negociación paritaria.

Esta distinción conceptual es de la mayor importancia. El convenio colectivo opera en un plano diferente al del consentimiento individual: mientras este se desenvuelve en la relación singular entre el responsable del tratamiento y cada interesado, el convenio lo hace en la relación colectiva entre la representación organizada de los trabajadores y el empleador. Y esa diferencia de planos no es accidental, sino constitutiva y ello en la medida en que el convenio está diseñado precisamente para actuar en ese ámbito en el que la voluntad individual es estructuralmente débil. Es, en cierta medida, la respuesta que el ordenamiento jurídico da a esa debilidad, como pone de manifiesto el artículo 88.1 del RGPD al reconocer expresamente los convenios colectivos como instrumento para establecer normas más específicas que garanticen la protección de los derechos y libertades en relación con el tratamiento de datos personales en el ámbito laboral ⁽³¹⁾. Reconocimiento que el Considerando 155 del RGPD extiende a los “convenios de empresa” ⁽³²⁾ y que permite a estos establecer “normas específicas relativas al tratamiento de datos personales de los trabajadores en el ámbito laboral”, se extiende a aspectos tan sensibles como “las condiciones en las que los datos personales en el contexto laboral pueden ser objeto de tratamiento sobre la base del consentimiento del trabajador”.

Esta función co-reguladora se manifiesta en tres planos complementarios. En primer lugar, en la posibilidad de definir finalidades legítimas en contextos sectoriales específicos, aportando criterios de necesidad vinculados a la organización del trabajo que difícilmente pueden ser captados por la norma legal general. En segundo lugar, en la introducción de condiciones de uso y límites operativos que inciden directamente en la proporcionalidad del tratamiento. En tercer lugar, en la incorporación de garantías adicionales que refuerzan la posición de los trabajadores, en línea con lo exigido por los artículos 88 y 9.2.b RGPD. Es precisamente en este tercer plano donde se sitúa una de

³¹ Tal y como se extrae del estudio de Halefom H. ABRAHA, “A pragmatic compromise? The role of Article 88 GDPR in upholding privacy in the workplace”, *International Data Privacy Law*, 2022, vol. 12, núm. 4, pp. 261-282, los convenios colectivos constituyen uno de los ejes centrales del artículo, tanto como instrumento habilitado por el artículo 88 RGPD para regular el tratamiento de datos en el ámbito laboral, como mecanismo que refleja la dimensión colectiva del Derecho del Trabajo que el RGPD, por su carácter marcadamente individualista, no logra acoger plenamente.

³² El referido Considerando 155 en su versión castellana se refiere a los “convenios de empresa”. La versión inglesa lo hace a los “*works agreements*”, mientras que la alemana y la francesa utilizan una expresión equivalente: “*Betriebsvereinbarungen*” y “*accords d’entreprise*”. Conceptos, en todo caso, bien diferenciados de “*Tarifverträge*” y “*conventions collectives*”. Ello plantea la duda de si, en el caso español, el referido concepto debe alcanzar también a los acuerdos de empresa en sentido estricto (tal y como son concebidos en el Derecho del Trabajo español, con un régimen jurídico propio y diferenciado del convenio colectivo estatutario). Con todo, resulta significativo que la STJUE de 19 de diciembre de 2024 (Asunto C-65/23, *K GmbH*) se refiera, precisamente, a un acuerdo de empresa como instrumento sobre cuya validez y compatibilidad con el RGPD se pronuncia el Tribunal, lo que apunta a que el ámbito de aplicación del artículo 88 del RGPD no se circunscribe a los convenios colectivos en sentido formal, sino que se extiende a otras modalidades de negociación colectiva de ámbito empresarial, con independencia de su configuración técnica en cada ordenamiento nacional.

las aportaciones más significativas de la negociación colectiva en materia de protección de datos: la que concierne al control del poder unilateral del empleador.

En ausencia de negociación colectiva, el empleador retiene una amplia discrecionalidad para decidir qué datos recaba, con qué finalidad, durante cuánto tiempo y a través de qué medios. Esa discrecionalidad está sometida, ciertamente, a los límites que impone el marco general de protección de datos –en particular los principios del artículo 5 del RGPD y las condiciones de licitud de los artículos 6 y 9 del mismo texto–, pero esos límites son, en gran medida, límites negativos. El convenio colectivo transforma esa situación de dos maneras complementarias. De un lado, convierte los principios generales en obligaciones concretas y verificables, dotando de contenido positivo a lo que de otro modo sería un marco de prohibiciones abstractas. De otro, desplaza la iniciativa regulatoria desde la decisión unilateral del empleador hacia el proceso de negociación, en el que los trabajadores, a través de sus representantes, tienen voz activa, conforme al artículo 91 de la LOPD, que reconoce la función del convenio como instrumento de densificación normativa de garantías adicionales de los derechos digitales en el ámbito laboral, en una intervención que no es meramente declarativa o programática, sino materialmente configuradora de los tratamientos.

Esta transformación proyecta un efecto democratizador sobre las condiciones de tratamiento de datos en la empresa, no porque la negociación colectiva sea un proceso intrínsecamente perfecto ni porque los representantes de los trabajadores encarnen siempre con plena fidelidad los intereses de la totalidad de los afectados, sino porque introduce un contrapeso institucional al poder del empleador que el consentimiento individual es, por su propia naturaleza, estructuralmente incapaz de proporcionar. Con todo, esta virtualidad no está exenta de límites, como advirtió el Tribunal de Justicia de la Unión Europea en la Sentencia de 19 de diciembre de 2024 (Asunto C-65/23, *K GmbH*), al señalar que las partes negociadoras no pueden alcanzar compromisos que, por razones económicas o de mera conveniencia, menoscaben indebidamente el objetivo cardinal del RGPD de garantizar, a un nivel elevado, la protección de las libertades y los derechos fundamentales de los trabajadores.

La afirmación del valor de la negociación colectiva como instrumento de protección de datos no puede, sin embargo, llevar a ignorar sus límites intrínsecos. El más importante de ellos es que el convenio colectivo no puede suprimir ni comprometer de manera irreversible los derechos individuales de los trabajadores en materia de protección de datos. Así lo ha confirmado la STS 12 de marzo de 2026 (Rº 55/2025), al concluir que las disposiciones convencionales no son autosuficientes ni pueden derogar ni excluir la aplicación de los artículos 5, 6 y 9 del RGPD, de modo que el convenio puede activar o estructurar el tratamiento, pero su licitud depende en todo caso de su integración en el sistema del Reglamento. El convenio puede definir condiciones, establecer marcos, introducir garantías –en los términos exigidos por el artículo 9.2.b) del RGPD para los datos de categorías especiales– y articular obligaciones. Pero no puede prestar, en nombre de los trabajadores, un consentimiento que estos no han dado individualmente para tratamientos que afectan de manera directa y personal a sus

derechos fundamentales, cuya indisponibilidad colectiva deriva, en el ordenamiento español, de su condición de derechos reconocidos en la Sección Primera del Capítulo Segundo del Título I de la CE. La dimensión colectiva del trabajo no disuelve la dimensión individual de la persona, el trabajador es siempre, al mismo tiempo, miembro de un colectivo y sujeto individual con derechos propios que no pueden ser objeto de disposición colectiva.

De este límite se extrae una consecuencia conceptual de primer orden y esta no es otra que la negociación colectiva no puede ser concebida como un mecanismo de externalización del consentimiento ni como una fórmula para abaratar el coste de obtener la aquiescencia de los trabajadores a tratamientos que de otro modo serían inaceptables. Si se utilizara con esa finalidad, no solo traicionaría el propósito que le es propio –el que le reconocen los artículos 88 y 9.2.b) del RGPD en cuanto instrumentos de co-regulación que concretan el contenido material de la legalidad del tratamiento–, sino que produciría el efecto contrario al que se pretende: en lugar de reforzar la protección de los trabajadores, la debilitaría al amparo de una apariencia de acuerdo. La legitimidad del convenio colectivo en este ámbito depende, en último término, de que responda genuinamente al interés de los trabajadores y no de que simplemente formalice la aquiescencia colectiva a tratamientos diseñados unilateralmente por el empleador. La distinción entre ambos supuestos no siempre es fácil de trazar en la práctica, pero es conceptualmente imprescindible; y su control, como ha establecido el TJUE en la citada Sentencia de 19 de diciembre de 2024, debe poder ejercerse sin restricción alguna sobre el respeto de todas las condiciones y límites establecidos por el RGPD para el tratamiento de datos personales.

4. El consentimiento colectivo ante la IA: una nueva dimensión de alta intensidad

El consentimiento ocupa un lugar central en el debate contemporáneo sobre la democracia en el trabajo. El Informe Final de la Comisión Internacional de Expertos de Alto Nivel sobre la Democracia en el Trabajo lo aborda como una categoría compleja y multidimensional que trasciende la mera formalidad jurídica para convertirse en el núcleo de una gobernanza empresarial genuinamente democrática. Esta dimensión se despliega en dos planos cualitativamente diferenciados: el del consentimiento genuinamente informado, articulado a través del derecho de consulta, y el del consentimiento con efecto de bloqueo, expresado mediante un derecho de veto.

En el primer plano, el artículo 64.5 del Estatuto de los Trabajadores (en adelante, ET) reconoce al comité de empresa el derecho a ser informado y consultado sobre todas las decisiones que pudieran provocar cambios relevantes en la organización del trabajo y en los contratos. La implantación de un sistema de IA que redefina procedimientos de evaluación, asignación de tareas o selección de personal encaja directamente en ese supuesto. Es, por definición, una decisión con impacto relevante sobre la organización. El contraste con el Derecho alemán resulta ilustrativo: el *Betriebsrat* debe ser informado

y consultado en la fase de planificación del uso de IA, de modo que sus propuestas y objeciones puedan incidir en la decisión antes de que esta se adopte ⁽³³⁾. El ordenamiento español, en cambio, corre el riesgo de relegar la consulta a un momento en que las decisiones están ya sustancialmente cerradas.

Ahora bien, ni la amplitud del objeto de la consulta ni la oportunidad de su ejercicio garantizan por sí solas la efectividad del consentimiento. Un consentimiento basado en información incomprensible no es un consentimiento libre, es una formalidad vacía que puede resultar más perjudicial que la ausencia de cualquier mecanismo participativo, porque crea una apariencia de legitimidad que encubre una ausencia real de control democrático. Para que el consentimiento sea genuinamente informado, el reconocimiento del derecho a contar con la asistencia de una persona experta en IA, a cargo de la empresa, no es un lujo accesorio, sino una condición de posibilidad ⁽³⁴⁾. Sin ese apoyo técnico, los representantes de los trabajadores se reducen, en la práctica, a ratificar lo que el empleador ha presentado como única opción disponible, reproduciendo en el plano tecnológico la misma asimetría estructural que el consentimiento individual era incapaz de superar en el de los datos personales.

El segundo plano es el más controvertido. La consulta, incluso cuando es previa e informada, no vincula al empleador. El Informe va más lejos y propone el reconocimiento de un derecho del comité de empresa a dar su consentimiento a la introducción de sistemas de IA en todos los asuntos que afecten al trabajo, con posibilidad de vetar todas las decisiones relacionadas con su despliegue ⁽³⁵⁾. En el plano europeo, esta dimensión se proyecta también sobre los Comités de Empresa Europeos. El Informe propone ampliar el papel del CEE para incluir un poder de codecisión en asuntos de IA, por cuanto la participación «debe ir más allá» de la mera información y consulta, requiriendo «derechos reales de toma de decisiones» sobre las vías tecnológicas que configuran directamente la vida laboral de los trabajadores ⁽³⁶⁾.

La propuesta, con todo su valor programático, no está exenta de objeciones. Un poder de veto de perímetro tan amplio aplica una respuesta uniforme a situaciones de riesgo cualitativamente distintas y reconfigura el poder de dirección sin la necesaria reformulación normativa de sus bases. La taxonomía de riesgos que ofrece el propio Reglamento de Inteligencia Artificial de la Unión Europea podría servir de referente para articular respuestas diferenciadas y proporcionales: un derecho de veto efectivo para los sistemas de alto riesgo que afectan a decisiones de empleo, y un derecho de consulta reforzada para los de riesgo menor. Esta modulación, que el Informe no desarrolla, per-

³³ Como explica Juan PEÑA MONCHO, *Los derechos de información y consulta frente a la dirección algorítmica del trabajo. Propuestas para impulsar la participación de las personas trabajadoras ante la IA*, Pamplona, Aranzadi, 2025, pp. 196-197.

³⁴ Informe Final de la Comisión Internacional de Expertos y Expertas de Alto Nivel sobre la Democracia en el Trabajo, Madrid, MITES, 2026, p. 343.

³⁵ Informe Final de la Comisión Internacional de Expertos y Expertas de Alto Nivel sobre la Democracia en el Trabajo, cit., p. 342.

³⁶ Informe Final de la Comisión Internacional de Expertos y Expertas de Alto Nivel sobre la Democracia en el Trabajo, cit., p. 454.

mitiría dotar al nuevo garantismo tecnológico de la precisión técnica y la coherencia sistemática que su conversión en Derecho positivo exige.

V. El principio de proporcionalidad preventiva o desde el diseño de la gestión algorítmica del trabajo

1. De la reactividad al diseño: la proporcionalidad como mandato preventivo

La irresistible transformación que acompaña al modelo tecnológico obliga a repensar las categorías jurídicas con las que el ordenamiento ha venido articulando la protección de los derechos fundamentales en el ámbito laboral. Cuando nos preguntamos por el alcance de esos derechos –intimidad, secreto de las comunicaciones, protección de datos– ante los sistemas algorítmicos y de IA, nos situamos ante una permanente confrontación entre las garantías fundamentales del trabajador y el interés empresarial en el establecimiento y uso de esas tecnologías. La técnica jurídica para resolver esa confrontación es, y seguirá siendo, el principio de proporcionalidad. Una herramienta que, descartando jerarquías formales entre los bienes en conflicto, permite alcanzar un ajuste material entre ellos mediante el análisis de idoneidad, necesidad y proporcionalidad en sentido estricto de la medida restrictiva ⁽³⁷⁾.

El fortalecimiento del principio de proporcionalidad ante los desafíos de la IA no requiere su sustitución, sino su transformación en varias dimensiones articuladas: temporal (de reactivo a preventivo), material (enriquecido con los principios del RGPD y con estándares técnicos verificables), colectiva (abierto a la participación de los trabajadores y sus representantes), y, en fin, sistémica (atento a los efectos agregados y no solo a los impactos individuales). Sin esa transformación plural, el principio de proporcionalidad corre el riesgo de convertirse en un instrumento de legitimación formal de decisiones tecnológicas adoptadas unilateralmente por quien detenta el poder, en lugar de operar como lo que siempre debería haber sido, esto es, un límite real y efectivo al ejercicio de ese poder.

El principio en su configuración clásica –diseñado para operar sobre medidas ya adoptadas, proyectando el juicio sobre algo que ya ha ocurrido– resulta estructuralmente insuficiente para contener los riesgos que generan los sistemas de IA. Estos riesgos no son plenamente visibles una vez desplegado el sistema pues se generan en el diseño mismo, en la estructura profunda del algoritmo y en la configuración de los criterios de decisión automatizada. Esperar a que el daño se produzca para proyectar sobre él el juicio de pro-

³⁷ Este trabajo contiene y continúa las reflexiones que inicié en Jesús R. MERCADER UGUINA, “El juicio de proporcionalidad en lo laboral: construcción, nuevas dimensiones y futuros escenarios”, *Anuario Coruñés de Derecho Comparado del Trabajo*, 2022, Volumen XIV, pp. 157-175. Y, más recientemente y en estrecha conexión con lo que aquí se dice en: “El principio de proporcionalidad como límite al control laboral basado en la inteligencia artificial”, *Trabajo y Derecho: Revista de actualidad y relaciones laborales*, 2024, nº 19 (monográfico Impacto Laboral de la Inteligencia Artificial).

porcionalidad equivale a aceptar, de facto, que ese daño es el punto de partida del sistema de garantías, y no el objeto de su prevención.

La respuesta a esta insuficiencia no consiste en abandonar el principio, sino en transformar su modalidad de operación de reactiva a preventiva, de ex post a ex ante, de correctiva a constitutiva. Esa transformación da lugar a lo que puede denominarse proporcionalidad por diseño: la obligación de evaluar, documentar y justificar la proporcionalidad de un sistema tecnológico antes de su implantación. Trasladado al ámbito laboral, este mandato obliga al empleador a demostrar, antes de implementar el sistema, que existe una finalidad legítima específica, que el sistema es idóneo para alcanzarla, que no existe un medio igualmente eficaz y menos invasivo para los derechos de los trabajadores, y que los datos tratados se reducen al mínimo imprescindible. La inversión de la carga argumentativa que esto supone es de la mayor importancia desde la perspectiva del garantismo laboral dado que es el empleador quien debe justificar la necesidad, no el trabajador quien debe probar la desproporción.

A la proporcionalidad por diseño debe añadirse, además, la exigencia de reversibilidad tecnológica como componente autónomo del juicio. El análisis clásico de proporcionalidad no incorpora una dimensión temporal de largo alcance y se limita a evaluar la medida en el momento de su adopción, pero no considera los efectos de su consolidación irreversible. En el contexto de la IA, esta omisión es particularmente relevante porque algunos sistemas, una vez implantados, generan dependencias organizativas y tecnológicas que hacen prácticamente imposible su remoción posterior. Incorporar la reversibilidad como criterio del juicio de proporcionalidad implica exigir que el empleador acredite, en la evaluación previa, que el sistema puede ser desactivado, modificado o sustituido sin consecuencias irreversibles sobre los derechos de los trabajadores, y reconocer que una medida que, aun siendo inicialmente proporcionada, genera dependencias tecnológicas irreversibles, introduce un factor de desproporción sobrevenida que debe ser valorado de manera continua durante toda la vida útil del sistema.

2. Los principios de protección de datos como contrafuertes del juicio de proporcionalidad

El principio clásico de proporcionalidad, en su formulación jurisprudencial consolidada, no es suficiente por sí solo para contener los riesgos que los sistemas de IA generan sobre los derechos fundamentales de los trabajadores ⁽³⁸⁾. La solución no pasa por sustituirlo, sino por enriquecerlo con los principios que la normativa de protección de datos ha desarrollado para operar específicamente en el entorno de los tratamientos masivos de datos personales. Una lectura integrada del principio de proporcionalidad y de los principios inspiradores del RGPD permite construir las bases de una proporcionalidad reforzada.

³⁸ Antonio ALVAREZ DEL CUVILLO, "Articulación entre la Ley y el juicio de proporcionalidad en la protección de los derechos fundamentales de las personas trabajadoras frente a la disrupción tecnológica", en AA.VV., *Derecho del Trabajo, cambios económicos y nueva sociedad*, Valencia, Tirant lo Blanch, 2025, pp. pp. 85-117.

El primero de los principios que actúa como refuerzo es el de transparencia. La opacidad en el conocimiento de los componentes esenciales de los sistemas digitales convierte en inútil el posterior juicio: sin transparencia no cabe proporcionalidad alguna. La exigencia de transparencia no es, en este contexto, una obligación de información de carácter accesorio, sino el presupuesto lógico del propio juicio de proporcionalidad. Un empleador que despliega un sistema de IA sin informar adecuadamente a los trabajadores y a sus representantes sobre su funcionamiento, sus criterios de decisión y sus efectos sobre las condiciones de trabajo no solo incumple un deber instrumental, sino que hace imposible el examen de si el sistema reúne las condiciones que ese juicio exige. Para que esa transparencia sea operativa, resulta necesario desarrollar estándares técnicos de explicabilidad que traduzcan la exigencia jurídica a parámetros verificables: qué nivel de comprensibilidad de las decisiones automatizadas es exigible en función de la intensidad del impacto sobre los trabajadores, y a través de qué mecanismos concretos debe satisfacerse esa exigencia.

El segundo principio es el de limitación de la finalidad. Los fines para los que son recabados los datos constituyen la pauta de la proporcionalidad en su tratamiento, puesto que la idoneidad, necesidad y proporcionalidad de este se miden en función del fin para el que se han obtenido. Proyectada sobre el juicio de proporcionalidad laboral, la limitación de la finalidad refuerza el juicio de justificación de la medida empresarial. La herramienta de IA debe ser utilizada exclusivamente para los fines específicos para los que fue diseñada e implantada, sin derivar hacia otros usos que no hayan sido sometidos a su propio juicio de proporcionalidad. Esta exigencia adquiere especial relevancia en el contexto de los sistemas de IA, cuya capacidad de aprendizaje y adaptación puede llevarlos, en ausencia de controles adecuados, a operar sobre finalidades que no fueron objeto de evaluación en el momento de su implantación.

El tercero es el de minimización, que viene a dotar de contenido más consistente a la exigencia de necesidad ⁽³⁹⁾. La minimización no es una simple limitación cuantitativa al volumen de datos que pueden ser recabados: es la expresión material del subprincipio de necesidad aplicado al tratamiento de datos, que obliga a elegir siempre aquella opción técnicamente viable que implique el menor grado de injerencia en los derechos fundamentales de los trabajadores. Esta exigencia adquiere una importancia capital cuando se toma en cuenta que la restricción en el uso de la información se contrapone a la lógica misma del Big Data, lo que obliga a interpretar de forma tendencialmente restrictiva el alcance de la limitación del derecho fundamental que el juicio de proporcionalidad debe valorar. Junto a la minimización, los principios de exactitud y limitación del plazo de conservación condicionan igualmente la proporcionalidad en

³⁹ Ello justificaría, como han expresado algunos autores, que el principio de minimización de datos no se conciba como “un test de proporcionalidad al uso, sino un principio de funcionamiento”, de modo que la aplicación de este principio lleve consigo “autorizar la recolección de datos dentro del contrato de trabajo, no cuando es necesario en el sentido de útil, sino solamente cuando es completamente imprescindible”. Ello llevaría consigo el recurso a sistemas “menos invasivos, como son la vigilancia humana –y no tecnológica–, Adrián TODOLI SIGNES, “Control tecnológico: una propuesta de aplicación del triple juicio de proporcionalidad conforme a la normativa europea de protección de datos” en *XXXII Congreso Anual de la Asociación Española de Derecho del Trabajo y de la Seguridad Social*, Madrid, MTES, 2022, pp. 241-242

sentido estricto, dado que datos inexactos o conservados más allá de lo necesario dejan de ser proporcionales, aunque la finalidad que justificó su obtención fuera legítima en su momento.

La operatividad de estos principios en entornos de alta complejidad técnica exige, por último, el desarrollo de indicadores sectoriales de proporcionalidad, elaborados mediante negociación colectiva sectorial o en el marco de organismos paritarios, que concreten qué umbrales de control, qué volúmenes de datos y qué finalidades son admisibles en cada tipo de actividad. Sin esa concreción técnica, el riesgo de que los principios operen como cláusulas retóricas sin contenido verificable es real y debe ser conjurado institucionalmente.

3. La evaluación de impacto relativa a la protección de datos y la evaluación de impacto en derechos fundamentales del artículo 27 RIA como expresión procedimental del principio de proporcionalidad en el entorno digital

La evaluación de impacto relativa a la protección de datos (EIPD) a la que se refiere el artículo 35 del RGPD ⁽⁴⁰⁾, no es un instrumento auxiliar del juicio de proporcionalidad, sino su expresión procedimental más genuina en el entorno digital, al constituir un proceso diseñado para describir el tratamiento, evaluar su necesidad y proporcionalidad, y gestionar los riesgos para los derechos y libertades de las personas físicas determinando las medidas para hacerles frente ⁽⁴¹⁾. La EIPD debe llevarse a cabo con carácter previo al tratamiento, lo que la convierte en una herramienta de apoyo a la toma de decisiones coherente con los principios de protección de datos desde el diseño y por defecto. Esta exigencia de anticipación resulta especialmente relevante en el contexto de los sistemas de inteligencia artificial desplegados en el ámbito laboral, dado que la obligación de realizar una EIPD es particularmente pertinente cuando se introduce una nueva tecnología de tratamiento de datos. Es precisamente sobre esta lógica anticipatoria donde el artículo 27 del RIA construye su mandato más ambicioso: la evaluación de impacto en materia de derechos fundamentales (EIDF), que los responsables del despliegue de sistemas de IA de alto riesgo deben realizar con carácter previo a su puesta en servicio, extendiendo el ámbito de análisis más allá de la protección de datos para abarcar el conjunto de derechos fundamentales que el sistema puede comprometer, incluidos los derechos colectivos de las personas trabajadoras.

La evaluación de impacto en protección de datos (EIPD) y la evaluación de impacto en derechos fundamentales (EIDF) constituyen instrumentos complementarios e indisolubles en el ámbito del despliegue de sistemas de IA. Ambas evaluaciones comparten una lógica de proporcionalidad que no se agota en un juicio binario: si el tratamiento

⁴⁰ Miguel RECIO GAYO, “Aproximación basada en el riesgo, evaluación de impacto relativa a la protección de datos personales y consulta previa a la autoridad de control”, en José Luis PIÑAR MAÑAS, *Reglamento General de Protección de Datos: Hacia un nuevo modelo europeo de privacidad*. Madrid, Reus, 2016, pp. 351 a 366

⁴¹ En el ámbito laboral, Jesús BAZ RODRÍGUEZ, *Privacidad y protección de datos de los trabajadores en el entorno digital*, Madrid, Wolters Kluwer, 2019, pp. 328-330.

no supera el test de proporcionalidad, deben adoptarse medidas correctoras, salvaguardas y mecanismos de protección que reduzcan el riesgo a un nivel aceptable. Tienen asimismo en común una exigente dimensión documental, orientada a garantizar la trazabilidad de las decisiones adoptadas y la rendición de cuentas, tanto interna como externa, si bien el artículo 27 RIA va más allá al imponer la notificación de los resultados de la EIDF a la autoridad de vigilancia del mercado, reforzando así el control público sobre el despliegue de la tecnología. Cuando el responsable del despliegue haya realizado previamente una EIPD conforme al artículo 35 del RGPD, la EIDF deberá complementarla y no sustituirla, consagrando normativamente la necesidad de una lectura sistemática e integrada de ambos marcos regulatorios.

No obstante, ambas evaluaciones presentan, en su configuración actual, una limitación estructural que condiciona su efectividad real como mecanismos de control y esta no es otra que quien examina la proporcionalidad del sistema es el mismo sujeto que lo ha diseñado o adquirido. Esta circunstancia genera un conflicto de interés inherente que debilita la objetividad del análisis y cuestiona la capacidad de ambos instrumentos para actuar como garantías independientes frente a los riesgos que pretenden controlar. Para corregir esa debilidad, resulta necesario articular mecanismos de auditoría algorítmica independiente que introduzcan un control externo hoy inexistente con carácter general. El fin de una auditoría es identificar o anticipar errores, riesgos o amenazas (actuales o potenciales) y ayudar a corregirlos ⁽⁴²⁾.

Esa auditoría debería operar en tres momentos: con carácter previo a la implantación, verificando que el sistema cumple los parámetros de proporcionalidad declarados por el empleador; de manera periódica durante su funcionamiento, dado que los sistemas de IA aprenden y evolucionan, pudiendo generar sesgos o efectos desproporcionados sobrevenidos; y con una orientación específica hacia la detección del sesgo algorítmico, orientada a identificar si el sistema genera impactos desproporcionados sobre grupos de trabajadores definidos por razón de género, edad, origen o cualquier otra característica protegida. La publicidad de los resultados de esa auditoría –al menos frente a los representantes de los trabajadores y la autoridad de control– es condición de su efectividad como mecanismo de garantía.

A ello debe añadirse una dimensión frecuentemente desatendida, la proporcionalidad transnacional. Los sistemas de IA que operan en entornos laborales son frecuentemente diseñados, entrenados y controlados en jurisdicciones distintas a aquella en que se aplican. Esta disociación entre el lugar de diseño y el lugar de impacto genera un vacío en el control de proporcionalidad que las normas nacionales y europeas no pueden colmar por sí solas. Resulta necesario, por ello, imponer al empleador que adquiere un sistema de IA desarrollado por un tercero el deber de verificar que ese sistema cumple los parámetros de proporcionalidad exigibles en el ordenamiento del lugar de aplicación, con

⁴² Jorge GARCÍA HERRERO, “¿Cómo auditar un algoritmo?”, entrada de 28 de febrero de 2017 en el blog <https://jorgegarciaherrero.com/auditar-algoritmo/> De este tema me he ocupado, en Jesús R. MERCADER UGUINA, *Algoritmos e inteligencia artificial en el derecho digital del trabajo*, Valencia, Tirant lo Blanch, 2022. También Adrián TODOLÍ SIGNED, *Algoritmos productivos y extractivos*, Pamplona, Aranzadi, 2023, pp. 90-92, que apuesta por la auditorías públicas.

independencia de dónde fue diseñado, estableciendo una responsabilidad compartida entre el desarrollador y el empleador-responsable del despliegue que impida la externalización de esa responsabilidad hacia el proveedor tecnológico.

VI. El principio de transparencia accionable en la gestión algorítmica

1. La transparencia como garantía estructural del nuevo garantismo tecnológico

El garantismo clásico edificó su sistema de protección sobre un conjunto de normas imperativas capaces de resistir la presión del desequilibrio contractual inherente a la relación laboral. La transparencia, en ese sistema, era un instrumento accesorio. El trabajador tenía derecho a conocer las condiciones de su contrato, pero el proceso mismo de dirección y control empresarial permanecía esencialmente opaco. La transformación digital ha invertido esa lógica de manera radical. Cuando las decisiones que afectan a los trabajadores son adoptadas o sustancialmente determinadas por sistemas algorítmicos, la transparencia deja de ser un complemento del sistema de garantías para convertirse en su presupuesto lógico.

La transparencia, sin llegar a configurarse como derecho fundamental, constituye una garantía esencial de esta nueva modernidad jurídica. Como explica Ferrajoli, “las garantías no son otra cosa que las técnicas previstas por el ordenamiento para reducir la distancia estructural entre normatividad y efectividad, y, por tanto, para posibilitar la máxima eficacia de los derechos fundamentales”. Por ello, continúa, “reflejan la diversa estructura de los derechos fundamentales para cuya tutela y satisfacción ha sido previstas”. En todo caso, “el garantismo de un sistema jurídico es una cuestión de grado que depende de la precisión de los vínculos positivos o negativos impuestos a los poderes por las normas constitucionales y por el sistema de garantías que aseguran una tasa más o menos elevada de eficacia de tales vínculos” (⁴³). Sin transparencia no hay posibilidad real de ejercer derechos frente a las decisiones algorítmicas; sin transparencia, la protección es meramente formal.

La transparencia en la IA se refiere, en sentido estricto, a la cualidad de un sistema de IA de poder ser enteramente interpretable y comprensible por todas las personas que lo crean e interactúan con él a lo largo de todo su ciclo de vida, de modo que sea posible entender el porqué de sus razonamientos y controlarlo adecuadamente (⁴⁴). El problema es

⁴³ Luigi FERRAJOLI, *Derechos y garantías. La Ley del más débil*, Madrid, Trotta, 1999, p. 25. En su contribución más reciente sobre el significado del garantismo (“El actual significado del garantismo”, *Teoría y Realidad Constitucional*, 2025, nº 55, p. 95), el propio Ferrajoli precisa el carácter activo de esa función de garantía. Las garantías son el “lado activo, hecho de obligaciones y prohibiciones”, de los derechos fundamentales, que consisten en expectativas pasivas. Esta distinción entre la dimensión pasiva del derecho (la expectativa) y la dimensión activa de la garantía (la obligación o la prohibición) tiene una traducción directa en el ámbito de la transparencia algorítmica. La transparencia no puede operar solo como expectativa del trabajador (el derecho a ser informado si lo solicita), sino que debe configurarse como obligación positiva del empleador, exigible con independencia de cualquier reclamación individual.

⁴⁴ Lilian EDWARDS y Michael VEALE, “Slave to the Algorithm? Why a “Right to an Explanation” is Probably

que muchos sistemas de IA modernos no cumplen esta condición. La distinción entre los llamados modelos de “caja negra” y los de “caja blanca” resulta aquí esencial. En los primeros, su funcionamiento interno no resulta explicable por sí mismo, de modo que se conoce la entrada y la salida, pero el proceso que media entre ambas es difícilmente comprensible incluso para los técnicos que los implementan. En los segundos, en cambio, los algoritmos empleados son más sencillos de entender y es posible interpretar cómo las características de entrada se transforman en la variable de salida o de destino, lo que hace su comportamiento más accesible al escrutinio: así sucede con la regresión lineal o los árboles de decisión.

La dificultad reside en que son precisamente los modelos de mayor complejidad, como las redes neuronales y los algoritmos de deep learning, los que ofrecen mayor precisión y resultan más aptos para casos de uso complejos, hasta el punto de que las representaciones necesarias para ilustrar su funcionamiento interno pueden llegar a ser tan complejas de comprender como los propios modelos, lo que genera una tensión estructural entre eficacia técnica y transparencia ⁽⁴⁵⁾.

La naturaleza de esta opacidad puede ilustrarse recurriendo a un referente proveniente de la física cuántica: la paradoja de Schrödinger ⁽⁴⁶⁾. Esta parte de un principio fundamental: mientras el sistema permanece cerrado y no es observado, el gato está simultáneamente vivo y muerto, en una superposición de estados indeterminados, sin que exista un “estado verdadero” accesible hasta que se produce la observación. Los modelos de caja negra comparten precisamente esta condición de opacidad estructural. A pesar del creciente uso de la IA, muchos de estos sistemas operan de formas que son opacas tanto para quienes los proveen como para quienes los despliegan y para quienes son afectados por su uso; en el complejo ámbito de los sistemas de IA, incluso los propios proveedores son con frecuencia incapaces de explicar las decisiones y resultados de los sistemas que han construido. Así, el interior de la caja negra se asemeja al interior de la caja de Schrödinger. Un espacio donde los procesos ocurren, pero cuya naturaleza exacta permanece inaccesible desde el exterior.

Esta opacidad se traduce en consecuencias materiales de primera magnitud en ámbitos concretos de la vida social y económica. La tensión entre eficacia técnica y transparencia se manifiesta con especial intensidad en el ámbito de las relaciones laborales. Los modelos de IA pueden descubrir nuevas correlaciones entre determinadas

Not the Remedy You Are Looking For”. *Duke Law & Technology Review*, 2017, nº 16, pp. 18-84. Lorenzo COTINO HUESO, Jorge CASTELLANOS CLARAMUNT (dirs.) *Transparencia y explicabilidad de la Inteligencia Artificial*, Valencia, Tirant lo Blanch, 2022. Lucía ORTIZ DE ZÁRATE ALCARAZO, “Explicabilidad (de la inteligencia artificial)”. *Eunomía. Revista en Cultura de la Legalidad*, 2022, nº. 22, pp. 328-344. Adrián TODOLÍ SIGNES, *Algoritmos productivos y extractivos*, cit., pp. 81-83. Eva BLAZQUEZ AGUDO, *El derecho de información como elemento estructural del principio de transparencia en las relaciones laborales*, Pamplona, Aranzadi, 2024.

⁴⁵ EUROPEAN DATA PROTECTION SUPERVISOR (EDPS), *TechDispatch on Explainable Artificial Intelligence (XAI)*, 2023, p. 6. En los modelos de “caja blanca” (*white box*) los algoritmos utilizados son sencillos de entender y es posible interpretar cómo las características de entrada se transforman en la variable de salida o de destino. Es posible identificar las características más importantes para predecir la variable objetivo, y estas características son comprensibles. Los modelos de “caja negra” (*black box*), en cambio, no son explicables por sí mismos. La falta de explicabilidad puede deberse a una ofuscación intencional por parte del diseñador del sistema o a la complejidad del modelo.

⁴⁶ Sobre el significado de la paradoja, David LINDLEY, *Incertidumbre*, Barcelona, Ariel, 2008, pp. 202-208.

características de entrada y tomar decisiones basadas en modelos altamente complejos que implican un gran número de parámetros en interacción, posiblemente millones, lo que hace difícil incluso para los expertos en IA comprender cómo se producen sus resultados. Un sistema de gestión de horarios o de asignación de rutas en una plataforma de reparto que utilice redes neuronales profundas puede optimizar la eficiencia operativa de forma extraordinaria, pero al mismo tiempo resultar completamente opaco para el repartidor que recibe instrucciones que considera perjudiciales para sus condiciones de trabajo.

Ante esta realidad, la IA Explicable (*Explainable Artificial Intelligence* o *XAI*) se ha configurado como la capacidad de los sistemas de IA de proporcionar explicaciones claras y comprensibles sobre sus acciones y decisiones, con el objetivo central de hacer comprensible para los seres humanos el comportamiento. De ello se deriva la exigencia de “idear formas de que los sistemas expliquen en profundidad sus decisiones o las abran al escrutinio”, lo que se ha convertido en un rompecabezas técnico fundamental para los investigadores de la materia. Existen, no obstante, “algunos indicios prometedores de que los modelos de IA podrían ser capaces de justificar sus resultados, aunque no de razonarlos en términos causales todavía”. En este sentido, se está llevando a cabo una importante labor en “el uso de estructuras simplificadas para explorar otras más complejas, incluso en la automatización del propio proceso de investigación de la estandarización: construir inteligencias artificiales que nos ayuden a contener la propia IA” ⁽⁴⁷⁾.

2. De la transparencia informativa a la transparencia accionable: el paradigma de la explicabilidad

La distinción fundamental en materia de transparencia no reside en el volumen de información divulgada, sino en su utilidad operativa para quien la recibe. El Big Data y sus aplicaciones mediante sistemas de IA plantean un desafío estructural al ordenamiento jurídico: la opacidad no es un defecto corregible de estos sistemas, sino, al menos en el estado actual del desarrollo tecnológico, una característica intrínseca ⁽⁴⁸⁾. Esta opacidad impide que pueda ofrecerse universalmente una explicación satisfactoria de cómo los datos de entrada se transforman en resultados de salida. Además, dificulta supervisar, identificar y probar posibles vulneraciones del Derecho, incluidas las disposiciones que protegen los derechos fundamentales. El cambio de paradigma que se opera con el RIA no es, por tanto, el de sustituir la opacidad por la transparencia total –empresa técnicamente imposible–, sino el de articular garantías jurídicas que hagan que la información disponible sea accionable para quienes son afectados por las decisiones algorítmicas.

⁴⁷ Los entrecomillados pertenecen a Mustafá SULEYMAN, Michel BHASKAR, *La ola que viene. Tecnología, poder y el gran dilema del siglo XXI*, Barcelona, Debate, 2025, p. 277.

⁴⁸ Michele FINC, *The EU Artificial Intelligence Act. A Commentary*, Oxford, University Press, 2026, p. 10.

La distinción entre transparencia, interpretabilidad y explicabilidad no es meramente académica, sino que tiene consecuencias jurídicas directas en la delimitación de las obligaciones que el RIA impone a proveedores y responsables del despliegue. Desde esta perspectiva, “la transparencia puede responder a la pregunta de “qué ocurrió” en el sistema; la explicabilidad puede responder a “cómo” se tomó la decisión; y la interpretabilidad puede responder a “por qué” se tomó esa decisión y cuál es su significado para el usuario”. Conviene advertir, sin embargo, que el RIA tiende a difuminar las fronteras entre estos conceptos, llegando a confundir sus significados e implicaciones, lo que genera incertidumbre en la delimitación práctica de las obligaciones normativas. La transparencia es una característica “pasiva” del modelo que permite acceder a información sobre el sistema; la explicabilidad es una característica “activa” orientada a generar explicaciones adaptadas a la audiencia destinataria.

Por su parte, la relación de implicación recíproca entre transparencia, explicabilidad y supervisión humana constituye el núcleo conceptual del sistema de garantías tecnológicas. Uno de los hallazgos más importantes del análisis del RIA es que la explicabilidad y la supervisión humana no son garantías independientes, son vasos comunicantes. Sin transparencia no hay supervisión efectiva, y sin supervisión efectiva no hay explicación significativa. La cadena se cierra pues el proveedor debe diseñar el sistema con suficiente transparencia, el responsable del despliegue debe ejercer la supervisión con el conocimiento que esa transparencia le proporciona y, cuando la decisión perjudica a alguien, esa misma cadena de comprensión es la que permite articular la explicación que exige el artículo 86 del RIA.

En cuanto a la titularidad de las garantías de transparencia, la información individual corresponde a cualquier persona trabajadora, la información colectiva a la representación de las personas trabajadoras, y la explicabilidad a la persona trabajadora individualmente considerada. En cuanto al momento de ejercicio, la información individual y colectiva opera *ex ante* —en la fase de diseño y definición de parámetros—, mientras que la explicabilidad opera *ex post*, vinculada a la decisión concreta que ha afectado al trabajador.

3. El derecho a la explicación individualizada: el artículo 86 del Reglamento de IA

La explicabilidad en IA se concentra en proporcionar razones claras y coherentes para predicciones o decisiones específicas del modelo. Aspira a responder preguntas del tipo “¿Por qué el sistema de IA realizó esta predicción concreta?” ofreciendo justificaciones o razones comprensibles para el ser humano sobre un resultado específico ⁽⁴⁹⁾. La explicabilidad requiere la interpretabilidad como elemento básico, pero también recurre a otros campos

⁴⁹ Me remito más ampliamente a Jesús R. MERCADER UGUINA, “*El derecho a explicación de decisiones tomadas individualmente*”, en la publicación que se realizará próximamente y que recogerá las ponencias del II Congreso Internacional. La IA en el mundo del trabajo La Directiva de Trabajo en Plataformas, el Reglamento General de Protección de Datos y el Reglamento de Inteligencia Artificial, Valencia, 25 de junio de 2026.

y ámbitos, como la interacción humano-máquina, el Derecho y la ética ⁽⁵⁰⁾. La explicabilidad es especialmente importante en aplicaciones críticas en las que están en juego vidas humanas o información sensible, ya que ayuda a usuarios, reguladores y partes interesadas a comprender el razonamiento que subyace a los resultados generados por la IA.

La ISO define la explicabilidad como la propiedad de un sistema de IA de expresar factores importantes que influyen en sus resultados de una manera comprensible para los humanos ⁽⁵¹⁾. La UNESCO (2021), señala que la explicabilidad “*hace referencia al hacer inteligible los resultados de los sistemas de IA, incluyendo la comprensibilidad de los datos, procesos y comportamientos de los distintos bloques algorítmicos y cómo cada uno de ellos contribuye al resultado del sistema, estando estrechamente relacionada con la transparencia y la trazabilidad*” ⁽⁵²⁾. El Grupo de Expertos de Alto Nivel en IA en el apartado 53 de sus directrices señala que “*la explicabilidad concierne a la capacidad de explicar tanto los procesos técnicos de un sistema de IA como las decisiones humanas asociadas*”. Añade que “*la explicabilidad técnica requiere que las decisiones que adopte un sistema de IA sean comprensibles para los seres humanos y estos tengan la posibilidad de rastrearlas*”. Para los sistemas de IA más opacos o de caja negra, precisa que debe acudir “*a medidas relacionadas con la explicabilidad (por ejemplo, la trazabilidad, la auditabilidad y la comunicación transparente sobre las prestaciones del sistema)*” ⁽⁵³⁾.

La explicabilidad traza fuertes puntos de conexión con la idea de motivación. Y ello en la medida en que motivar una decisión es expresar sus razones y por eso es necesario obligar al que la toma a exponerlas y justificarlas. Pero la motivación también permite saber cómo y por qué se ha tomado una decisión. La exigencia de motivación tiene por finalidad asegurar el conocimiento de la parte sobre las razones de hecho y de derecho que justifican una determinada decisión, como garantía de la exclusión de la arbitrariedad, y como instrumento para posibilitar su impugnación ⁽⁵⁴⁾. La motivación debe justificar y rendir cuentas ⁽⁵⁵⁾.

La distinción entre transparencia e interpretabilidad, de un lado, y explicabilidad, del otro, es sintetizada con precisión por la Guía española: la transparencia es una

⁵⁰ EDPS, *TechDispatch on Explainable Artificial Intelligence (XAI)*, cit., p. 5.

⁵¹ La norma ISO/IEC 42001: 2023, “Tecnología de la información -- Inteligencia artificial -- Sistema de gestión”, establece que los sistemas de gestión de IA deben garantizar niveles adecuados de transparencia y explicabilidad, dos principios esenciales para asegurar confianza, control y responsabilidad en el uso de modelos de inteligencia artificial. La explicabilidad exige que los resultados generados por la IA puedan ser entendidos y justificados. La norma subraya que el nivel de explicabilidad debe ser proporcional al impacto del sistema, especialmente cuando afecta a derechos, seguridad o cumplimiento normativo.

⁵² La referencia se extrae de Lucía ORTIZ DE ZÁRATE ALCARAZO, “Explicabilidad (de la inteligencia artificial)”, cit., p. 334.

⁵³ GRUPO INDEPENDIENTE DE EXPERTOS DE ALTO NIVEL SOBRE INTELIGENCIA ARTIFICIAL, *Directrices éticas para una IA fiable*, Bruselas, Comisión Europea, 2019, p. 16.

⁵⁴ En el caso de la sentencia la idea es clara, la fundamentación de contenido de una sentencia, como reiteradamente ha recordado el Tribunal Constitucional, “permite cumplir la doble finalidad de la exigencia de motivación, exteriorizar de un lado, el fundamento de la decisión adoptada, y permitir, de otro, su eventual control jurisdiccional mediante el efectivo ejercicio de los derechos. SSTC 199/1991, de 28 de octubre, FJ. 3; 208/1993, de 28 de junio, FJ. 1.

⁵⁵ Sobre la exigencia de motivación como “manera de dar cuentas”, José CALVO GONZÁLEZ, *Derecho y narración. Materiales para un teoría y crítica narrativista del Derecho*, Barcelona, Ariel, 1996, Capítulo I y II.

característica “pasiva” del modelo, mientras que la explicabilidad es una característica “activa” orientada a generar explicaciones adaptadas a la audiencia destinataria. Aplicar medidas de transparencia sobre el sistema de IA permite entender su funcionamiento e interpretar sus resultados. Como consecuencia, y de manera bidireccional, la transparencia y la supervisión humana están completamente relacionadas, ya que para que el sistema pueda ser supervisado por personas es fundamental que el sistema sea transparente ⁽⁵⁶⁾.

El artículo 86 del RIA reconoce que toda persona afectada por una decisión adoptada ⁽⁵⁷⁾ por el responsable del despliegue sobre la base de los resultados de un sistema de IA de alto riesgo ⁽⁵⁸⁾ que produzca efectos jurídicos o la afecte de manera considerable (y que considere que le causa un perjuicio para su salud, su seguridad o sus derechos fundamentales), tendrá derecho a obtener del responsable del despliegue explicaciones claras y significativas sobre el papel del sistema de IA en el proceso decisorio y los principales elementos de la decisión adoptada. En virtud de este derecho, quien haya sido rechazado para un puesto de trabajo, a quien se haya denegado una ayuda social o a quien un sistema de IA haya clasificado como de alto riesgo, podrá exigir que se le explique qué papel desempeñó dicho sistema y qué factores determinaron el resultado.

Conviene precisar la naturaleza técnica de la explicación que el artículo 86 del RIA demanda, pues no toda explicación sobre un sistema de IA satisface su exigencia. Cuando el sistema opera como una “caja negra” y las explicaciones se generan ex post, estas pueden ser globales (proporcionan una comprensión general del comportamiento del modelo en su conjunto, aplicable a todos los casos) o locales (se centran en esclarecer por qué el sistema adoptó una predicción o decisión particular en un caso concreto: “por qué este sistema tomó esta decisión concreta sobre mí”) ⁽⁵⁹⁾. El derecho que consagra el artículo 86 del RIA es, en rigor, un derecho a la explicación local e individualizada: no a comprender cómo funciona el sistema en general, sino a recibir una explicación referida a los factores que determinaron el resultado específico del interesado. Solo la explicación local permite al afectado ejercer sus derechos de manera efectiva y al supervisor humano controlar de manera sustantiva la decisión concreta, que es precisamente la exigencia que el legislador europeo ha querido garantizar.

⁵⁶ SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL, *Guía 8. Transparencia y provisión de información a los usuarios. Reglamento Europeo de Inteligencia Artificial*. Gobierno de España, 2024.

⁵⁷ El concepto de “decisión” no está definido. Teniendo en cuenta que este concepto también se utiliza en el artículo 22 del RGPD, podemos recurrir a la jurisprudencia pertinente, que subraya que dicho concepto debe interpretarse de manera amplia y que en él pueden incluirse determinadas tareas reparadoras. En el asunto SCHUFA, la Sentencia del tribunal de justicia (Sala Primera) de 7 de diciembre de 2023 sostuvo que el establecimiento automatizado, por parte de una agencia de información crediticia, de un valor de probabilidad basado en datos personales relativos a una persona y referentes a su capacidad para cumplir con sus compromisos de pago en el futuro constituye una decisión basada exclusivamente en un tratamiento automatizado.

⁵⁸ La RIA tampoco define el concepto de “resultado de un sistema de IA”. Sin embargo, cabe destacar que el considerando 171 establece que solo están sujetas al artículo 86, apartado 1, aquellas decisiones «basadas principalmente en el resultado de determinados sistemas de IA de alto riesgo».

⁵⁹ En la distinción que efectúa EDPS, *TechDispatch on Explainable Artificial Intelligence (XAI)*, cit.,

VII. El principio de reserva de humanidad real en la gestión automatizada del trabajo

1. El problema central: la presencia sin capacidad de decisión

La “reserva de humanidad”, como se la ha calificado ⁽⁶⁰⁾, la idea del humano al mando (*human-in-the-loop*, *human-on-the-loop*, o *human-in-command*) ⁽⁶¹⁾ constituye un soporte fundamental en el diseño de los nuevos sistemas de IA. La doctrina ha advertido desde el principio sobre el riesgo de que la exigencia de supervisión humana quedase vaciada de contenido mediante una participación meramente formal. La práctica demuestra que la supervisión humana con frecuencia no resulta eficaz por diversas razones: los supervisores pueden carecer de las competencias necesarias o tener incentivos contrapuestos que comprometan su juicio, deficiencias que plantean obstáculos significativos a la hora de implementar una supervisión humana efectiva ⁽⁶²⁾.

Como ha señalado parte de la doctrina, “los seres humanos no son un parche normativo o de diseño que pueda insertarse al azar como solución a problemas que, en realidad, tienen que ver con la forma en que está estructurado el sistema en su conjunto” ⁽⁶³⁾. El artículo 14.1 del RIA establece que “los sistemas de IA de alto riesgo se diseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas”. El Grupo de Trabajo del Artículo 29 ya había señalado en términos inequívocos que el responsable del tratamiento no puede obviar las disposiciones del artículo 22 del RGPD inventándose una participación humana, y que, para ser considerada como tal, el responsable del tratamiento debe garantizar que cualquier supervisión de la decisión sea significativa, en vez de ser únicamente un gesto simbólico, y debe llevarse a cabo por parte de una persona autorizada y competente para modificar la decisión. Los apartados 3 y 4 del artículo 14 ofrecen ejemplos de medidas eficaces de supervisión humana y subrayan que estas dependen del contexto, correspondiendo al proveedor determinar cuáles son adecuadas y proporcionadas en función del sistema de IA, su finalidad prevista y el estado actual de la técnica.

A medida que los algoritmos se integran en más y más procesos decisorios, los responsables políticos necesitan orientación más precisa sobre cómo utilizar el Derecho para fomentar una toma de decisiones híbrida que resulte verdaderamente productiva. El riesgo que se denuncia con mayor precisión es el de la llamada trampa MABA-MABA

⁶⁰ Juli PONCE SOLÉ, “Reserva de humanidad y supervisión humana de la inteligencia artificial”, *Cronista del Estado Social y Democrático de Derecho*, 2020, nº 100, pp. 58-67.

⁶¹ José VIDA FERNÁNDEZ, “Los retos de la regulación de la inteligencia artificial: algunas aportaciones desde la perspectiva europea”, en Tomás DE LA QUADRA-SALCEDO, José Luis. PIÑAR MAÑAS, (Dir.) *Sociedad Digital y Derecho*. Madrid. BOE, 2018, p. 218.

⁶² Guillermo LAZCOZ MORATINOS, “La vigilancia o supervisión humana en el artículo 14 de Reglamento de inteligencia artificial: ¿un mero requisito obligatorio para los sistemas de alto riesgo?”, en Lorenzo COTINO HUESO, Pere SIMON CASTELLANO (Dir.), *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea*, Pamplona, Aranzadi, 2024, pp. 681-699.

⁶³ Rebecca CROOTOF, Margot E KAMINSKI y William NICOLSON PRICE II, “*Humans in the Loop*”, *Vanderbilt Law Review*, 2023, 429, disponible en <https://scholar.law.colorado.edu/faculty-articles/1594>.

(*Men Are Better At / Machines Are Better At*, “*Los hombres son mejores en... / Las máquinas son mejores en...*”). Los responsables políticos a menudo asumen que añadir un ser humano a un sistema automatizado dará como resultado lo mejor de ambos mundos⁽⁶⁴⁾. Existe una seductora simplicidad en este enfoque de “inserta un humano”. Pero este planteamiento se basa en una premisa intuitiva, errónea y peligrosa cual es que los sistemas hombre-máquina representan siempre lo mejor de los dos y no introducen nuevos problemas propios. En el contexto laboral, insertar a una persona en el proceso de validación de una decisión algorítmica no garantiza en absoluto que esa persona pueda ni esté en condiciones reales de controlar, corregir o revertir el resultado del sistema.

Por ello, la doctrina más avanzada ha propuesto que resultaría más apropiado hablar de “experto en el bucle” (*expert in the loop*) que de “humano en el bucle” (*human in the loop*), en la medida en que la mera presencia formal de una persona es condición necesaria, pero en ningún caso suficiente para que la supervisión sea genuinamente efectiva⁽⁶⁵⁾. El artículo 14 del RIA define con precisión las condiciones que debe reunir la supervisión para ser genuinamente efectiva. El supervisor debe entender las capacidades y limitaciones del sistema y vigilar su funcionamiento; ser consciente del sesgo de automatización; interpretar correctamente los resultados; poder decidir no usar el sistema o descartar, invalidar o revertir sus salidas; e intervenir o detener el sistema mediante, por ejemplo, un botón de parada.

El mandato institucional que de ello se deriva es claro y no es otro que el que exige que los responsables del despliegue encomendarán la supervisión humana a personas físicas que tengan la competencia, la formación y la autoridad necesarias para ejercerla de manera real. La supervisión no es una función que pueda encomendarse a cualquier persona. Dicha función requiere cualificación técnica, formación específica sobre los riesgos del sistema y autoridad real para oponerse a sus recomendaciones sin sufrir consecuencias adversas. La supervisión es real cuando el supervisor puede decir “no” al algoritmo. El considerando 73 del RIA estipula que “las personas físicas a las que se haya asignado la supervisión humana cuenten con la competencia, la formación y la autoridad necesarias para desempeñar dicha función”. Este requisito específico del contexto se ajusta al mandato sobre conocimientos en materia de IA del artículo 4, que exige a los proveedores y a los responsables del despliegue garantizar que su personal cuente con los conocimientos adecuados en materia de IA. Si bien se trata de una obligación de medios y no de resultados, se convierte en una obligación de resultados en lo que respecta a las personas físicas que cumplen las obligaciones derivadas del artículo 14 RIA.

⁶⁴ Durante más de setenta años, una visión simplista pero problemática de los sistemas humano-máquina ha sido la de asignar tareas en función de lo que los “hombres hacen mejor” frente a lo que las “máquinas hacen mejor”. La lista de Fitts original de 1951 identificó, por ejemplo, que las máquinas son mejores en tareas repetitivas y rutinarias, mientras que los humanos son mejores en improvisar. La trampa MABA-MABA es el error regulatorio de creer que la mera presencia humana en un proceso de decisión algorítmica garantiza automáticamente un resultado mejor, sin considerar que el sistema híbrido resultante genera dinámicas y riesgos propios que deben ser específicamente diseñados, regulados y anticipados.

⁶⁵ Doria ONITIU, “The limits of explainability & human oversight in the EU Commission’s proposal for the Regulation on AI- a critical approach focusing on medical diagnostic systems”, *Information & Communications Technology Law*, 2023, n° 32, citado por Michele FINC, *The EU Artificial Intelligence Act. A Commentary*, cit., p.303.

El hecho de que la negociación colectiva comience a incorporar expresas obligaciones en esta materia es un indudable signo de la importancia de esta cuestión ⁽⁶⁶⁾.

2. El sistema integrado de garantías: transparencia, supervisión y explicación como triángulo normativo

El análisis de las garantías individuales no puede concluir sin una visión de conjunto que subraye su carácter sistémico. El derecho a la explicación no opera de forma aislada, sino que se inserta en una estructura de garantías interconectadas cuya eficacia depende de la articulación coherente de sus distintos elementos. La transparencia en el diseño del sistema (artículo 13 RIA), la supervisión humana efectiva (artículo 14 RIA), la evaluación de impacto en derechos fundamentales (artículo 27 RIA), la obligación de información al afectado (Considerando 93 RIA) y las vías de reclamación y recurso (artículos 85 y 99 RIA) conforman un ciclo de tutela en el que cada pieza refuerza a las demás. La quiebra de cualquiera de estos eslabones compromete la efectividad del conjunto: sin transparencia no hay supervisión efectiva, sin supervisión no hay explicación significativa, y sin explicación no hay tutela real frente a las decisiones algorítmicas.

El artículo 86 del RIA no puede comprenderse de forma aislada de los principios que vertebran el conjunto del Reglamento. En sistemas de IA de alto riesgo, el artículo 13 del RIA exige que se diseñen y desarrollen con un nivel de transparencia suficiente para que los responsables del despliegue interpreten y usen correctamente los resultados de salida. La relación entre el artículo 13 y el artículo 86 del RIA es, por tanto, instrumental: la transparencia prevista en el artículo 13 opera como precondition técnica que hace posible la explicación individualizada del artículo 86. Un sistema diseñado sin suficiente transparencia no puede, en la práctica, generar las explicaciones que el artículo 86 exige.

La representación normativa de este sistema integrado puede sintetizarse en lo que cabe denominar un “triángulo normativo”: la transparencia del artículo 13 del RIA como precondition técnica de la explicabilidad; la supervisión humana efectiva del artículo 14 del RIA como presupuesto de eficacia de la explicabilidad; y la explicación significativa del artículo 86 del RIA como expresión final de ese sistema de garantías ante el trabajador afectado por la decisión.

La supervisión colectiva de los sistemas de IA requiere, para ser efectiva, el mismo presupuesto que la supervisión individual: la transparencia. Decía Susan Sontag que “la transparencia supone experimentar la luminosidad del objeto en sí, de las cosas tal como son” ⁽⁶⁷⁾. Sin información sobre el funcionamiento del sistema –sus criterios, sus datos

⁶⁶ Roberto FERNÁNDEZ VILLARINO, “El derecho ante la inteligencia artificial. A propósito de lo dispuesto en el art. 80.5 del Convenio Colectivo Nacional de la Banca”. *Trabajo, Persona, Derecho, Mercado: Revista de Estudios sobre Ciencias del Trabajo y Protección Social*, 2021, 4, pp. 141-163. Más recientemente la disposición adicional cuarta del V Convenio Colectivo Estatal del Sector de Radiodifusión Comercial Sonora o el artículo 10 del XXI Convenio colectivo general de la industria química han venido a incorporar nuevas e interesantes reglas en esta materia.

⁶⁷ Susan SONTAG, “Contra la interpretación”, en *Contra la interpretación y otros ensayos*, Barcelona, Seix Barral, 1984, p. 26.

de entrenamiento, sus parámetros de decisión—, la representación de los trabajadores no puede ejercer una función real de contrapeso. El RIA parte de una complementariedad entre las Autoridades de Vigilancia del Mercado y las Autoridades que garantizan la protección de los trabajadores. En España, la Inspección de Trabajo y Seguridad Social continuará verificando que el empleador cumpla sus obligaciones de seguridad y salud. La AESIA supervisará que los sistemas de IA utilizados como producto cumplen los requisitos técnicos del RIA y en todo su ciclo de vida para evitar sesgos y discriminaciones. Ambas entidades actuarán de forma coordinada mediante mecanismos de comunicación e intercambio de información.

La transparencia, así concebida en su doble dimensión individual y colectiva, no es solo el presupuesto lógico del sistema de garantías tecnológicas: es también la condición de posibilidad de una gobernanza democrática de la IA en el ámbito laboral. Sin ella, el poder algorítmico del empleador se ejerce sin control ni contrapeso. Con ella, ese poder queda sometido al escrutinio de quienes lo padecen y de las instituciones diseñadas para protegerlos. Esa es, en definitiva, la función que el nuevo garantismo tecnológico le asigna a la transparencia. No un derecho más en el catálogo, sino la garantía de que todos los demás derechos pueden ser, en la práctica, ejercidos.

VIII. Conclusión: la continuidad del proyecto garantista bajo nuevas formas

El Derecho del Trabajo no ha llegado hasta aquí por accidente histórico, ni ha sobrevivido a cada gran transformación productiva por inercia institucional. Lo ha hecho porque su razón de ser, corregir el desequilibrio estructural de poder entre quien trabaja y quien organiza el trabajo, es una constante que ninguna revolución tecnológica ha podido eliminar, aunque sí ha modificado profundamente las formas en que ese desequilibrio se manifiesta y, por tanto, las herramientas con las que el ordenamiento jurídico debe hacerle frente. El ordenamiento no renunció a su función tuitiva y compensadora sino que encontró formas más complejas, más articuladas y adecuadas de realizarla. Esta lógica de continuidad funcional a través del cambio formal es la clave hermenéutica que permite comprender el momento presente sin caer ni en el catastrofismo de quienes proclaman la muerte del Derecho del Trabajo ni en la complacencia de quienes subestiman la profundidad del desafío que plantea la revolución digital y algorítmica.

Hoy asistimos a un proceso análogo al de aquellas grandes transiciones históricas. Los valores que el ordenamiento laboral debe proyectar sobre el nuevo territorio del desequilibrio de poder no son valores nuevos; son los valores constitucionales de siempre, que ahora deben operar sobre escenarios radicalmente distintos. El algoritmo que decide una contratación, el sistema de inteligencia artificial que evalúa el rendimiento, la plataforma que disciplina al trabajador sin empleador visible, la vigilancia digital que penetra en espacios antes impermeables al control empresarial: todos ellos son expresiones contemporáneas del mismo problema estructural que el Derecho del Trabajo lleva más de un siglo intentando gobernar.

La respuesta a ese desafío tiene un nombre a nuestro juicio: garantismo tecnológico. Y el garantismo tecnológico no es otra cosa que el constitucionalismo de siempre enfrentado a un problema nuevo. Sus exigencias son conocidas en su sustancia, aunque inéditas en su forma. Esa aspiración, como ha señalado Ferrajoli al proyectar el garantismo hacia sus exigencias contemporáneas, requiere hoy una dimensión que desborda los límites de los ordenamientos nacionales. La tercera fase histórica del garantismo –el garantismo global⁽⁶⁸⁾– es, en su concepción, la respuesta necesaria a la divergencia entre la vigencia de los derechos y su efectividad cuando los poderes que los amenazan operan a escala transnacional. El nuevo garantismo tecnológico laboral que aquí se ha analizado comparte esa lógica expansiva. Muestra de ello son el Reglamento de Inteligencia Artificial, el Convenio Marco del Consejo de Europa y el RGPD que resultan, precisamente, los instrumentos a través de los cuales el constitucionalismo y el garantismo están ensayando esa expansión en el territorio del trabajo digital.

La historia del Derecho del Trabajo es, en sí misma, la demostración de que la disciplina puede reinventarse sin perder su identidad. Ante la revolución industrial construyó los instrumentos tuitivos de primera generación. Ante la maduración del sistema de relaciones colectivas, los transformó en instituciones más sofisticadas y plurales. Ante la revolución digital y algorítmica, está construyendo un nuevo garantismo. El constitucionalismo digital y el garantismo tecnológico son los nombres contemporáneos de una aspiración que nació con el propio Derecho del Trabajo y que, mientras exista desequilibrio de poder en las relaciones laborales, no puede ni debe abandonarse.

⁶⁸ Luigi FERRAJOLI, “El actual significado del garantismo”, cit., pp. 89-100.