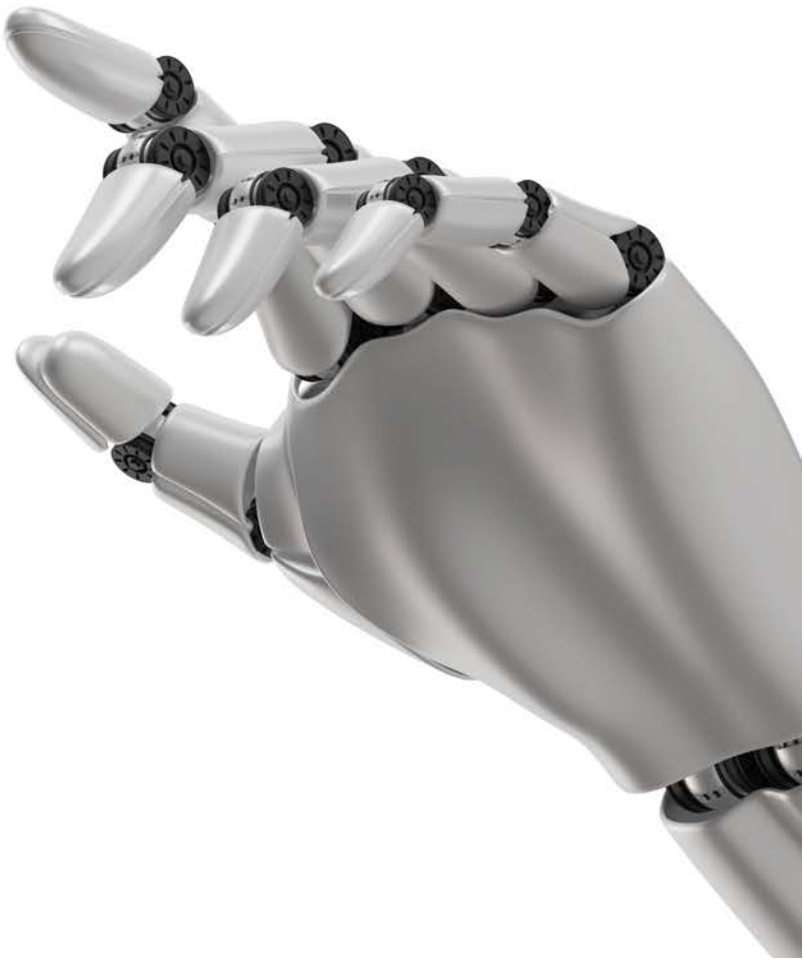
LIBRO BLANCO

+IGUAL

UN ENFOQUE PRÁCTICO, ÉTICO Y NORMATIVO PARA DESARROLLAR UN ESTÁNDAR DE CERTIFICACIÓN DE IA EN EL ÁMBITO LABORAL



Propuesta de actuación sobre el impacto de los sesgos para garantizar la equidad y la inclusión desde la Administración Pública y las empresas

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA
MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030

Plan de Recuperación, Transformación y Resiliencia

Comunidad de Madrid
Dirección General de Evaluación, Calidad e Innovación
CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Primera edición: julio 2025
ISBN: 978-84-09-73740-6
Copyright © **2025, Comunicación de Valor Añadido S.L.U.**
Información y contacto: Marisa Cruzado, coordinadora de IA+Igual.
mcruzado@cvalora.com
Coaturores: José Luis Poveda, Efrén Díaz, Marco Tulio Daza y Ana Cantalapiedra (DATAI - Universidad de Navarra)

Aviso legal y reserva de derechos
Todos los derechos están reservados. Queda estrictamente prohibida, sin autorización escrita del titular del copyright, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, incluidos la reprografía y el tratamiento informático, conforme a las sanciones establecidas en el ordenamiento jurídico vigente. Los contenidos de esta publicación son responsabilidad exclusiva del titular del copyright, por lo que la administración competente no asume responsabilidad alguna sobre dichos contenidos. Para solicitar la autorización de su reproducción total o parcial escribir al email: cva@cvalora.com

Abstract

El Libro Blanco del Proyecto IA+Igual explora las dimensiones éticas, sociales y legales de la integración de la inteligencia artificial en los procesos de Recursos Humanos. Financiado a través de la Consejería XX por el programa de la UE Next Generation, este documento destaca cómo los sesgos cognitivos y sistémicos integrados en los sistemas de IA pueden reforzar la desigualdad, la exclusión y la discriminación en el mercado laboral. Propone un modelo de gobernanza integral para el uso responsable de algoritmos en los sistemas de alto riesgo, como lo son los usados en RR. HH., centrado en la transparencia, la equidad y la supervisión humana. Las principales recomendaciones incluyen auditorías algorítmicas, protocolos de mitigación de sesgos y alineación con la Ley de IA de la UE. Está dirigido a todas aquellas personas -sean responsables políticos, organizaciones y profesionales de RR. HH.- que jueguen un papel en el desarrollo de sistemas de IA que defiendan la inclusión, la diversidad y los derechos fundamentales a la vez que optimiza la gestión del talento.

Índice

PRÓLOGO.....I
INTRODUCCIÓN..... VI

1. INTELIGENCIA ARTIFICIAL EN EL ÁMBITO LABORAL..... 1
1.1 Entendiendo la IA: definición, evolución y transformación en Recursos Humanos... 1
1.2 Beneficios y retos en la gestión de personas 7
1.3 Gobernanza del dato como pilar de una IA responsable e inclusiva.....23
1.4 IA aplicada a diferentes procesos en RR. HH.....31
1.5 Impacto de la IA en materia de diversidad e inclusión..... 43

2. MARCO LEGAL PARA IMPLANTAR LA IA EN EL ÁMBITO LABORAL .. 51
2.1 Marco legal en Europa..... 51
2.2 Marco legal de España..... 63
2.3 Marco legal global e internacional..... 79
2.4. Cuestiones jurídicas y derechos fundamentales en el ámbito laboral..... 84
2.5 Consideraciones éticas y legales..... 87

3. INVESTIGACIÓN EMPÍRICA DE IA EN RR. HH......106
3.1 Procesos de análisis de algoritmos.....106
3.2 Casos de uso analizados..... 132
3.3 Conclusiones empíricas.....151

4. RECOMENDACIONES PARA LA CERTIFICACIÓN DE SISTEMAS DE IA APLICADOS A RR. HH...... 153
4.1 Introducción y contexto.....153
4.2 Fundamentos para un sello de calidad “ética” en IA..... 154
4.3 Arquitectura del sello de calidad por niveles..... 156
4.4 Elementos clave del sistema de certificación algorítmica.....158
4.5 Recomendaciones para la implantación progresiva..... 160
4.6 Mecanismos de verificación y seguimiento..... 163
4.7 Consideraciones éticas y contextuales..... 165
4.8 Rol de las administraciones públicas en la promoción de la certificación ética en IA..... 166
4.9 Conclusiones..... 167
Apéndice del Capítulo 4.....169

5. RETOS FUTUROS Y CONCLUSIONES..... 178
5.1 El nuevo mercado laboral.....178
5.2 Escalabilidad de las soluciones planteadas..... 179
5.3 Retos legales futuros en un contexto de IA avanzada.....179
5.4 Conclusiones finales.....180

ANEXO.....184
GLOSARIO.....231
BIBLIOGRAFÍA.....252

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

PRÓLOGO

La universalidad de los sesgos

Un sesgo cognitivo es una interpretación errónea sistemática de la información disponible que ejerce influencia en la manera de procesar los pensamientos, emitir juicios y tomar decisiones. El concepto de sesgo cognitivo fue introducido por los psicólogos israelíes Kahneman y Tversky en 1972. Como hemos comentado, el cerebro ejecuta diariamente millones de procesos mentales y los sesgos son los atajos que toma para ser más eficiente. Atajos que influyen en la forma en cómo vemos el mundo y que están determinados por implicaciones culturales, influencia social, motivaciones emocionales o éticas, reducción de la información, o distorsiones en la recuperación de los recuerdos y la memoria, entre muchos otros. Los sesgos son, por tanto, algo inherente al ser humano.

La forma en que percibimos todo lo que nos rodea responde a la interpretación que nuestro cerebro –el de cada uno de nosotros– hace de ella. Y esa interpretación se define por herencia genética y también por el proceso de aprendizaje que llevamos a cabo de manera natural desde el momento en que nacemos.

Los sesgos nos ayudan a interpretar la realidad para que nos sintamos a gusto estemos alerta. Lo conocido, lo que se parece a nosotros, conforma nuestro espacio de confort. Lo que es diferente, lo que no reconocemos o actúa de manera distinta, nos genera rechazo.

Esa forma sesgada de ver la realidad representa para cada uno de nosotros nuestra realidad. Este hecho biológico subyace detrás de lo que en la sociedad actual reconocemos como estereotipos y prejuicios; y alimenta las desigualdades, la discriminación y la injusticia social. Pero su impacto, va mucho más allá. Cuando llevamos a cabo investigaciones, desarrollamos tecnología o definimos protocolos de cualquier tipo, los sesgos limitan nuestra visión y nos hacen trabajar en función del prototipo que más se asemeja a nosotros. Esta situación ha sido documentada en numerosos análisis sobre tecnología y sesgo.

En una economía global como la actual, los sesgos son cada vez menos una ayuda y más un problema porque la diversidad forma parte de nuestro día a día y, en un entorno multicultural, los prototipos dejan de ser eficientes.

Algunos sectores, como el Farmacéutico o el Automovilístico, llevan décadas trabajando para evitar el impacto negativo de los sesgos en sus respectivas industrias. Vender en un mercado global precisa analizar cuál es el efecto de un medicamento en personas de una y otra raza; o cuáles son los síntomas de una enfermedad en función del sexo o la edad. Para comercializar un coche, hay que tener en cuenta los gustos de hombres y mujeres, y las tendencias sociales en uno y otro lado del mundo. El prototipo de hombre blanco, de 1,80 metros de altura y 80 kg. de peso ha dejado de ser eficiente para muchas compañías del mundo occidental.

La cuestión es que, el nacimiento de Internet, y luego la llegada de los dispositivos móviles, las conexiones de alta velocidad y la ampliación de la capacidad de almacenaje de datos en la mal llamada nube, han supuesto una generación exponencial de información que se ha almacenado sin supervisión, filtros o garantías de veracidad. Toda esa información online, derivada de una interpretación sesgada de la realidad, es la fuente de entrenamiento de los algoritmos que soportan la IA.

La irrupción en 2024 de la IA Generativa se alimenta de toda esa información y supone un reto para todos los países y sus respectivas economías.

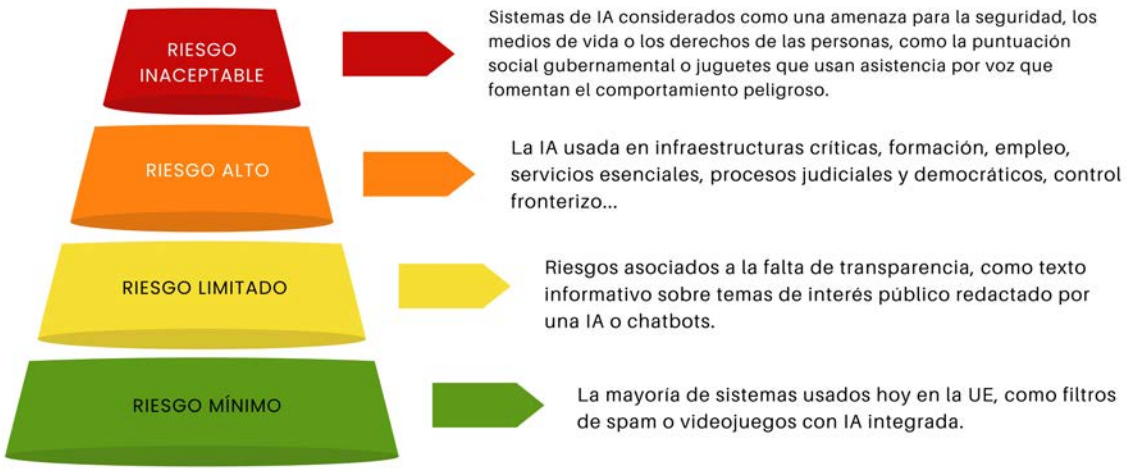
No cabe duda de su potencial dinamizador y de las oportunidades de desarrollo y crecimiento que su aplicación pone a disposición de las personas. Ni tampoco que es una tecnología cuyo uso transversal llega a todos los ámbitos, convirtiéndola no en una revolución tecnológica sino en una transformación social que está cambiando la forma en que nos relacionamos, trabajamos, viajamos, nos educamos o nos organizamos. Pero los riesgos también están claros, sobre todo, desde el punto de vista ético.

En este sentido, podemos identificar tres grandes líneas de trabajo globales. La primera de ellas tiene como objetivo el control sociopolítico. La IA se utiliza para la supervisión y clasificación de las personas con el objetivo de perpetuar un sistema político concreto. La segunda parte del control socioeconómico como objetivo. También se clasifica a las personas, pero con el fin último de fomentar el consumo. Detrás de este modelo están los grandes desarrolladores tecnológicos.

Ninguno de estos modelos superaría un sistema de control ético. No hay reglas ni normas, nadie cuestiona los para qué de su uso. Y en muchas ocasiones, el modelo que subyace es una caja negra que no permite entender su funcionamiento.

La tercera línea de trabajo pone a la persona en el centro del algoritmo y se corresponde con las líneas de trabajo de los gobiernos de la UE principalmente. Marcada por el riesgo de perder el tren de la competitividad, pone en marcha un proceso legislativo con el objetivo de innovar y generar tecnología soportada en IA, teniendo como hoja de ruta que dichos desarrollos deben estar al servicio de las personas y garantizar, por encima de todo, el respeto de los ODS, los ESG y los fundamentos de los Derechos Humanos. En el verano de 2024, se aprobó el primer Reglamento europeo de Inteligencia Artificial (RIA) que empezó a entrar en vigor en enero de 2025 y cuyo punto de partida es la clasificación de los usos de la IA, en función de sus niveles de riesgo.

Gráfico 1. Niveles de Riesgo de los modelos de IA



Fuente: elaboración propia.

Dicho reglamento recomienda a los estados miembro que empiecen a desarrollar mecanismos de supervisión de la IA. Las certificaciones se posicionan como una herramienta eficiente, que ha dado buenos resultados en otros ámbitos y que implica el desarrollo de normas y la aplicación de sanciones para garantizar su cumplimiento. Este documento responde a esta recomendación.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

El Libro Blanco que tienen en sus manos es el resultado del Proyecto IA+Igual, financiado con fondos Next Generation a través del área de Innovación Social de la Comunidad de Madrid, y con el impulso del Plan de Resiliencia y Recuperación del Ministerio de Derechos Sociales y Agenda 2030 del Gobierno de España. Gira en torno a tres ejes de actuación:

Eje 1.- Promoción de los derechos de las personas, la autonomía y la inclusión social, apartado 1 (Promover la autonomía personal y la inclusión social).

Eje 3.- Transformación digital, apartados 1 (Promover los derechos y la inclusión digitales) y 3 (Promover la reducción de la brecha digital).

Eje 5.- Inserción sociolaboral, apartados 1 (Mejorar la empleabilidad de las personas en situación de vulnerabilidad) y 2 (Acompañar procesos de inserción sociolaboral de personas vulnerables mediante itinerarios personalizados).

Desarrollado entre el 1 de julio de 2023 y el 30 de junio de 2025, ha analizado de forma empírica diversas herramientas soportadas en IA y aplicadas a procesos de Recursos Humanos (RR. HH.) con el objetivo de identificar el impacto de los sesgos, conscientes o inconscientes, en sus resultados.

El objetivo es claro: sentar las bases de un futuro modelo de certificación de herramientas de IA aplicadas a procesos del mercado laboral y de recursos humanos, para garantizar que se mantienen los avances conseguidos en materia de inclusión, equidad y diversidad. De esta forma, el proyecto contribuye a alcanzar los objetivos del Componente 22 del Plan de choque para el Economía de los Cuidados y Refuerzo de las políticas de igualdad e inclusión del Plan de Recuperación, Transformación y Resiliencia, que establece el marco necesario para impulsar el servicio asistencial español y favorecer un avance definitivo hacia la igualdad y la inclusión. En este ámbito, el plan hace especial referencia a aquellos proyectos que garanticen "que los servicios sociales y de empleo sean capaces de proporcionar un apoyo efectivo"; que "aumenten la tasa de empleo de las personas beneficiarias del Ingreso Mínimo Vital mediante el desarrollo del incentivo al empleo y se mejorará la coordinación y la gobernanza entre los agentes relevantes para la política de inclusión" y, más concreto aún, que garanticen "que los servicios sociales y de empleo sean capaces de proporcionar un apoyo efectivo" a los colectivos más vulnerables.

Las empresas que han colaborado con el proyecto han puesto a disposición del equipo de analistas e investigadores las herramientas de IA que estaban utilizando para prevenir fuga de talento, seleccionar personal, analizar el engagement de los empleados con la cultura corporativa o planificar el desarrollo de carrera. También hemos accedido a diseños del modelo de Gobernanza del dato o de redefinición de la descripción de los puestos de trabajo en el ámbito de los servicios públicos de empleo. Todo ello, nos ha ofrecido una visión amplia sobre: el estado de la cuestión; los retos y frenos a los que se enfrentan tanto las empresas desarrolladoras como las usuarias; la complejidad que implica la toma de decisiones en las organizaciones a la hora de implantar herramientas de IA en sus procesos de recursos humanos; al gran desconocimiento generalizado sobre al alcance de la IA, el impacto de los sesgos y las consecuencias de un uso inapropiado de la IA en el ámbito laboral.

Nos ha ayudado a constatar el riesgo que supone la irrupción de la IA en el mercado laboral para las personas de colectivos vulnerables o con riesgo de exclusión y los retos a los que se enfrentan las organizaciones para garantizar que no sé retrocede en materia de equidad, igualdad, diversidad e inclusión. El Reglamento General de Protección de Datos impide que las características de estos colectivos estén presentes en los datos laborales y, por tanto, son un agujero negro para la IA que no entiende la riqueza diversa de nuestra sociedad como ya se ha señalado en distintos estudios sobre justicia algorítmica. La interpretación sesgada de

la realidad llega a la inteligencia artificial a través de los datos con los que se entrena y de los que aprende.

El trabajo desarrollado por los profesionales de IA+Igual arroja una visión empírica del impacto que suponen los sesgos en el ámbito de la inclusión socio laboral. El equipo de este proyecto de innovación propone líneas de actuación para garantizar, no solo un uso ético, explicable y transparente de la inteligencia artificial en el ámbito laboral, sino para sentar las bases de un entrono inclusivo que sitúe a las personas en el centro del algoritmo; y contribuir a construir una sociedad diversa, equitativa e inclusiva más productiva, eficiente, rentable y competitiva.

Cuando este proyecto empezó hace dos años, no existía el RIA, ni tampoco normas ISO para la certificación de algoritmos. Empezaba a generalizarse el uso de ChatGPT, Copilot, y muy pocos sabían lo que son los agentes de IA. Por eso, este trabajo se enmarca en un entorno dinámico, incierto y cambiante que supone un auténtico reto para los gobiernos y la sociedad en su conjunto. Esperamos que, al menos, el contenido de este Libro Blanco arroje algo de luz sobre el mejor camino a seguir para el futuro del ser humano.

Agradecimientos

Este proyecto se inicia una mañana del mes de mayo de 2019. Aquel día coincidí con la entonces directora general de la Mujer de la Comunidad de Madrid, Laura Ruiz de Galarreta, en un evento sobre IA aplicada a movilidad y me propuso que desarrollara un proyecto que uniera dos conceptos: "tecnología e igualdad". Esta propuesta tan sencilla fue el germen de IA+Igual, un proyecto innovador, multidisciplinar y colaborativo que culmina con la elaboración de este Libro Blanco.

En estos seis años, muchas han sido las personas que se han ido uniendo al proyecto y a las que debo agradecer su confianza casi ciega en mí.

Como palanca de arranque, a Concepción Dancausa que aceptó recibirnos en 2021 para que le presentáramos el embrión del proyecto, cuando ocupaba el puesto de consejera de Políticas Sociales de la Comunidad de Madrid. Siempre gracias.

La palanca de anclaje ha sido la persona que ha colaborado directamente en la parte técnica y que completó la visión 360º de IA+Igual: Félix Villar. Y por extensión, el equipo de Altia, en especial: Jordi Carreras, José Luis Jorge, Manuel Muñoz, Elena Nikolaeva, Fernando Quintanilla y Amanda Figueredo.

Y como motor de desarrollo, al equipo del Instituto de Ciencias del Dato e Inteligencia Artificial (DATAI) de la Universidad de Navarra. En especial, a Iván Cordón, que ha sido mi apoyo y cuya coordinación del equipo multidisciplinar de académicos e investigadores ha dado tan magnifico resultado.

También muestro mi gratitud a los coautores: José Luis Poveda, Efrén Díaz, Marco Tulio Daza y Ana Cantalapiedra; y a los miembros de DATAI Jesús López Fidalgo, Ester Rodríguez-Losada Torres, Alberto García y Juan Carlos Hernández Peña (quien ha intervenido a través de conversaciones). Siempre gracias.

No puedo olvidar la generosidad de José Domínguez, José Luis Fernández, Yolanda Gutiérrez, Olga Herrero, Anabel Muñoz, Jesús Mercader, Javier Moscoso, Ambrosio Nguema y Ana Valera, todos miembros de nuestro Consejo Asesor, que nos han acompañado desde 2023 supervisando, valorando y aportando su tiempo y conocimiento al proyecto.

A nuestros embajadores Carlos Cabañas, Javier

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Campelo y Estela Martín, que siempre responden a mi llamada de auxilio y a los colaboradores Guillermo Butler, María Isadora Romero López, Joaquín García Onrubia, Virginia Carcedo Illera, Borja García Rubio, Cristina Martín, Elena Martín de Diego y Manuel Zaera. No cabe ninguna duda de su grandeza humana y profesional.

Debo agradecer también la implicación del departamento de Organización, Comp. & Beneficios y People Analytics del Corte Inglés, y a los equipos de Dathum, Aggity, Alma de Carraovejas, Atisa, Enagas, Optare Solutions, Fundación Once, Xunta de Galicia, Hospitales San Juan de Dios; a la Fundación a la Par y a la Asociación Autismo Madrid.

Tengo muy presente a la Comunidad de Madrid. A Ana Dávila-Ponce de León Municio, consejera de Familia, Juventud y Asuntos Sociales que cogió la batuta de Dancausa, a Pablo Gómez-Tavira Gómez-Tavira, su Viceconsejero, y a Óscar López Santos, Director General de Evaluación, Calidad e Innovación, así como al equipo del Área de Innovación, que nos ha sufrido y disfrutado a partes iguales. Gracias a todos ellos.

Asimismo, me siento agradecida a Patricia Reyes, directora general de la Mujer, y a Mª del Mar de Vicente, subdirectora General de Igualdad de Oportunidades. Siempre están ahí.

Y, por último, a mi equipo: Asun Velasco y Anabel Cuevas han sido mis manos, mis piernas, mis ojos y muchas veces mi cabeza a lo largo de estos años. Sin su trabajo, la dimensión de este proyecto no habría sido posible. Ni tampoco la calidad de este Libro Blanco que tienen en sus manos. Siempre, gracias.

Marisa Cruzado
Coordinadora del Proyecto IA+Igual y del Libro Blanco

INTRODUCCIÓN

La inteligencia artificial (IA) y las tecnologías basadas en el análisis de datos están generando transformaciones profundas en la economía global. Aunque su impacto total aún está por definirse, diversos estudios coinciden en que su potencial es significativo. Según Goldman Sachs, la IA generativa podría elevar el PIB mundial en un 7 % en la próxima década. Otras estimaciones, como la de PwC, sitúan esta subida en un 15 % hacia 2035, crecimiento que sería similar al que experimentó la economía global en el XIX cuando se industrializó (Goldman Sachs, 2023; PwC, 2025).

Estas tecnologías emergentes están impulsando la competitividad y la eficiencia en sectores como la Salud, las Finanzas o el Transporte, y ya están transformando la toma de decisiones estratégicas a través del análisis masivo de datos. Su alcance, sin embargo, no se limita a los sectores tradicionalmente tecnológicos: se espera que su influencia se extienda con fuerza a ámbitos como la gestión del talento y el trabajo humano, como analiza Rafael Andersson Lipcsey en su estudio sobre los efectos transformadores de la IA en la economía internacional (Lipcsey, 2023).

En este contexto, los Recursos Humanos juegan un papel clave. Su función en la atracción, desarrollo y retención del talento se ha vuelto aún más crítica en un entorno laboral marcado por la incertidumbre, la aceleración tecnológica y la necesidad de adaptación constante. Adoptar la IA en este ámbito no solo permite optimizar procesos como la selección o el desarrollo profesional, sino que también abre la puerta a experiencias más personalizadas para los empleados y decisiones más informadas basadas en datos (El País, 2024).

Sin embargo, la adopción de estas tecnologías no está exenta de desafíos. A pesar de los avances, muchas organizaciones siguen enfrentando una escasez de talento con competencias digitales, una brecha que se agrava por la globalización y la diversidad creciente en los entornos de trabajo. Según datos recientes, más del 75 % de las empresas tiene dificultades para cubrir sus vacantes con perfiles adecuados (ORH, 2024; KPMG, 2023).

La IA, bien integrada, puede ayudar a enfrentar estos retos. Permite automatizar tareas repetitivas, detectar patrones ocultos y liberar tiempo para funciones más estratégicas. Además, contribuye a promover entornos más inclusivos y equitativos, reduciendo posibles sesgos en los procesos de selección o evaluación del desempeño. Como muestra la tesis de Alexandre Peretó Rovira (Peretó, 2023), el impacto de la IA no solo afecta a la eficiencia económica, sino también a los marcos institucionales, normativos y sociales que configuran el futuro del empleo.

Este Libro Blanco explora cómo la inteligencia artificial está redefiniendo el ámbito de Recursos Humanos, desde las oportunidades que ofrece hasta las barreras éticas, legales y operativas que aún persisten. El objetivo es proporcionar una base rigurosa, realista y útil para los distintos agentes implicados en la gestión del talento.

1. Objetivo del Libro Blanco

El objetivo principal de este documento es fomentar el entendimiento y la implementación de la IA en los procesos de gestión de personas, con especial atención a sus implicaciones éticas, legales y

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

organizativas. Se busca explorar cómo la inteligencia artificial redefinir prácticas clave como la atracción de talento, la selección, el desarrollo profesional y la personalización de la experiencia del empleado, aportando beneficios tangibles en términos de eficiencia, toma de decisiones y bienestar laboral.

Partiendo de la siguiente definición de algoritmo –"Conjunto de instrucciones que sigue un sistema informático para ejecutar una tarea o tomar una decisión"–, propone estrategias para garantizar la transparencia y la equidad en el diseño y uso de algoritmo; al mismo tiempo que se abordan riesgos como los sesgos y las asimetrías en el acceso a oportunidades. Se presta especial atención a la protección de la diversidad y a la inclusión de colectivos tradicionalmente excluidos, con el fin de promover una gestión del talento más justa, responsable y respetuosa con los derechos fundamentales. Esto incluye a personas con discapacidad, mayores de 45 años, migrantes, mujeres en sectores con baja representación y otros grupos con barreras estructurales en el acceso y permanencia en el empleo.

Este Libro Blanco también ofrece orientaciones prácticas para auditar los sistemas de IA en RR, HH. desde una triple perspectiva: técnica, legal y ética. A través del análisis de casos reales, se ilustran los impactos transformadores de estas tecnologías en el entorno laboral, así como las tendencias emergentes en su adopción. Todo ello se acompaña de una revisión crítica de las normativas vigentes en materia de privacidad de datos y cumplimiento regulatorio.

En coherencia con este propósito, el análisis y las propuestas recogidas en este Libro Blanco se han articulado siguiendo los ejes estratégicos establecidos por el proyecto IA+Igual. Estos ejes integran distintas dimensiones, técnica, ética, legal y social de la inteligencia artificial aplicada a los recursos humanos, y han servido de marco metodológico durante todo el proceso de elaboración. Su presencia asegura que cada propuesta responda tanto a criterios de innovación como a principios de inclusión, equidad y responsabilidad institucional.

Finalmente, el documento reflexiona sobre los desafíos futuros, entre ellos la necesidad de que las soluciones de IA sean escalables, adaptables y sostenibles en un entorno de trabajo dinámico, interconectado y en permanente evolución.

Integración de ejes estratégicos sociales, digitales y laborales

La elaboración del presente Libro Blanco ha seguido como referencia transversal tres ejes estratégicos fundamentales, alineados con políticas públicas orientadas a la inclusión social, la transformación digital y la inserción sociolaboral. Estos ejes han sido concebidos como principios estructurantes que no solo han guiado la selección de contenidos, sino que también sustentan las buenas prácticas, las recomendaciones metodológicas y las propuestas de certificación que aquí se recogen.

Eje 1: Promoción de los derechos de las personas, la autonomía y la inclusión social

Este eje responde a la necesidad de garantizar el acceso al empleo como vía de autonomía personal y de inclusión efectiva, especialmente para colectivos en situación de vulnerabilidad: personas con discapacidad, víctimas de violencia de género o personas en riesgo de exclusión social. Desde esta perspectiva, se pone el foco en la implementación de sistemas de selección de personal objetivos y libres de sesgos, que valoren únicamente la idoneidad profesional. El diseño metodológico del proyecto asume este principio como un compromiso ético y operativo con los derechos humanos y la justicia social.

Eje 3: Transformación digital

La transición hacia una economía digital exige asegurar el ejercicio pleno de los derechos digitales, así como la reducción de las desigualdades en el acceso y uso de las tecnologías. Este eje orienta el despliegue de una estrategia de sensibilización en torno a la inteligencia artificial y su aplicación en los procesos de recursos humanos, haciendo uso de tecnologías innovadoras como la gamificación o entornos virtuales inmersivos. El proyecto se dirige especialmente a población joven y sectores desfavorecidos, promoviendo una digitalización inclusiva que no profundice las brechas existentes, sino que contribuya activamente a cerrarlas.

Eje 5: Inserción sociolaboral

La mejora de la empleabilidad y el acompañamiento de itinerarios personalizados de inserción sociolaboral constituyen el núcleo de este eje, que se dirige especialmente a migrantes, personas ex tuteladas, personas en situación de discapacidad o mujeres víctimas de violencia. La propuesta parte de la premisa de que los sesgos algorítmicos en los sistemas de IA pueden representar una barrera silenciosa pero persistente para estos colectivos. Por ello, se aboga por herramientas de evaluación y selección basadas en algoritmos auditables y éticamente diseñados, capaces de mejorar las oportunidades laborales de quienes tradicionalmente han sido excluidos del mercado de trabajo. La fase piloto del proyecto permitirá validar estos enfoques en colaboración con empresas comprometidas con la equidad algorítmica.

Estos ejes, además de responder a las directrices europeas en materia de supervisión algorítmica, están alineados con la Agenda 2030 de la UNESCO, en particular con los Objetivos de Desarrollo Sostenible (ODS) n.º 5 (Igualdad de género), n.º 8 (Trabajo decente y crecimiento económico) y n.º 10 (Reducción de las desigualdades). Asimismo, se encuentran en sintonía con los objetivos estratégicos de la Comunidad de Madrid en el ámbito de Asuntos Sociales, consolidando así una perspectiva multiescalar e interinstitucional del proyecto.

1.1 ¿A quién va dirigido?

En primer lugar, se dirige a representantes institucionales, tanto del sector público como privado, que participan en la elaboración de políticas laborales, en el diseño de marcos regulatorios o en el impulso de estrategias de digitalización inclusiva. Esto incluye a responsables sindicales, expertos en sostenibilidad y ética organizacional, y agentes sociales comprometidos con el futuro del trabajo.

También se orienta a profesionales que intervienen directamente en el diseño y uso de estas tecnologías: responsables de RR. HH., especialistas en tecnología aplicada a la gestión del talento, equipos legales y de cumplimiento normativo. Todos ellos deben comprender el funcionamiento, los beneficios y los riesgos asociados al uso de la IA en procesos que afectan directamente a las personas.

Además, el documento puede resultar de utilidad para investigadores, analistas y estudiantes que deseen explorar los impactos económicos, sociales y organizacionales de la IA en el mundo laboral, así como para docentes que buscan incorporar esta temática en programas de formación avanzada.

Por último, contribuye a que la implementación de la IA en RR. HH beneficie también a quienes históricamente han estado menos representados en los entornos organizativos. Por ello, su enfoque promueve una transformación tecnológica alineada con los principios de inclusión, equidad y diversidad.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

En definitiva, esperamos que resulte de interés para todos los actores implicados en la aplicación, regulación, supervisión o análisis de la inteligencia artificial en el ámbito laboral.

1.2 Contenido y alcance del documento

Este Libro Blanco ofrece un análisis riguroso y accesible sobre el uso de la inteligencia artificial en el ámbito del mercado laboral y los Recursos Humanos. A través de un enfoque multidisciplinar, examina cómo estas tecnologías están transformando los procesos de gestión del talento, y qué condiciones deben darse para que dicha transformación sea eficaz, ética y alineada con el marco normativo vigente.

El contenido se organiza en torno a cuatro pilares principales:

1. Fundamentos y contexto: se presentan las bases técnicas, sociales y económicas que explican el auge de la IA en el mundo del trabajo, así como los marcos legales y estratégicos que guían su aplicación en Europa y en España.

2. Procesos de RR. HH. y casos de uso: se analiza cómo la IA se integra en los distintos procesos de gestión de personas en las organizaciones, desde las etapas del ciclo de vida del empleado —selección, onboarding, desarrollo, evaluación, desvinculación— hasta otros ámbitos clave, como la administración de personal (nómina, gestión documental), la cultura corporativa o la experiencia del empleado. Se incluyen ejemplos prácticos y tendencias emergentes.

3. Riesgos, ética y gobernanza: se abordan los desafíos relacionados con el sesgo algorítmico, la equidad, la transparencia y la auditoría de sistemas, así como las responsabilidades de las organizaciones en la gestión de estas herramientas.

4. Marco práctico de orientación: se ofrecen recomendaciones para intermediadores sociolaborales, representantes de los trabajadores, responsables de RRHH, diseñadores de soluciones, auditores y legisladores, con el objetivo de facilitar una adopción informada, segura y responsable de la IA.

El alcance del documento se centra en organizaciones que aplican o prevén aplicar soluciones de IA en procesos relacionados con personas, ya sea mediante herramientas propias o a través de soluciones de terceros. No se abordan en profundidad los aspectos técnicos del desarrollo algorítmico ni las arquitecturas específicas de software, salvo cuando resultan necesarios para ilustrar implicaciones prácticas o riesgos.

A lo largo del documento, se presta especial atención al papel que puede desempeñar la IA para reducir brechas de acceso al empleo y visibilizar talento en colectivos históricamente infrarrepresentados. Esta perspectiva transversal se aborda tanto en los casos de uso como en las recomendaciones normativas y éticas, reforzando el compromiso con una gestión del talento inclusiva y socialmente responsable.

2. Metodología empleada

La elaboración rigurosa y participativa del *Libro Blanco Enfoque práctico, ético y normativo para desarrollar un estándar de certificación de IA en el ámbito laboral* ha seguido una metodología estructurada que está diseñada para garantizar la solidez técnica, la pluralidad de perspectivas y el

alineamiento con el marco normativo vigente. El proceso se ha extendido a lo largo de 16 meses, entre enero de 2024 y junio de 2025, y ha involucrado diversas fases, agentes y herramientas.

Fases del proceso

El trabajo comenzó con una sesión de diseño inicial en la que se definió un primer índice de contenidos de carácter general. A partir de ahí, se articuló una metodología compuesta por las siguientes fases principales:

- **Recogida de información**, tanto externa al proyecto como generada en su propio desarrollo.
- **Redefinición del índice** en función del análisis de la información recopilada.
- **Redacción de contenidos por epígrafes**, a cargo de diferentes autores expertos en la materia.
- **Revisión interna** por parte del Consejo Asesor.
- **Revisión externa** por profesionales especializados.
- **Petición y validación de colaboraciones,** garantizando calidad y coherencia de las aportaciones.
- **Corrección y edición** en dos rondas, unificando estilo y estructura.
- **Maquetación y diseño** de infografías.
- **Validación final** del documento.

Actores implicados

El proceso ha contado con la participación de una diversa red de actores procedentes del ámbito empresarial, institucional y académico, así como con la supervisión de un Consejo Asesor multidisciplinar que ha actuado como órgano consultivo y de validación en todo el proceso.

Fuentes y evidencias

Se ha desarrollado un enfoque basado en la combinación de diversas fuentes:

- Revisión documental y análisis de tendencias.
- Aportaciones de expertos a través de entrevistas individuales y artículos originales.
- Referencia constante al marco legal español y europeo relativo a la inteligencia artificial.
- Análisis de casos de uso reales, centrados en la implementación de algoritmos de IA en organizaciones.

Enfoque metodológico

El proyecto se ha sustentado en un enfoque de trabajo multidisciplinar y colaborativo, reflejando así la pluralidad de conocimientos necesarios para abordar la aplicación de la IA en la gestión de personas. La selección de casos de uso ha permitido observar la realidad práctica de la tecnología, integrando perspectivas técnicas, éticas, jurídicas y organizacionales.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Uso de herramientas digitales e inteligencia artificial

Como parte del compromiso con la innovación, se ha hecho un uso responsable y eficiente de herramientas de inteligencia artificial:

- **ChatGPT** ha sido utilizado en la redacción de borradores preliminares y en la elaboración del glosario del documento.
- También se ha empleado para **apoyar la búsqueda bibliográfica** y sistematizar contenidos.

Garantías éticas y de calidad

Con el objetivo de asegurar la calidad, confidencialidad y representatividad del contenido:

- Todos los agentes participantes han firmado acuerdos de confidencialidad.
- Cada epígrafe ha sido revisado por expertos en la materia correspondiente.
- Los textos han sido validados a través de una plataforma colaborativa integrada por el Consejo Asesor y por profesionales externos con experiencia contrastada.

1. INTELIGENCIA ARTIFICIAL EN EL ÁMBITO LABORAL

La inteligencia artificial es una de las tecnologías más influyentes en la transformación digital del mundo laboral (Brynjolfsson & McAfee, 2014). En los últimos años, su capacidad para procesar grandes volúmenes de datos, identificar patrones complejos y automatizar decisiones ha redefinido no solo cómo se gestionan los procesos operativos, sino también cómo se diseñan las estrategias de talento en las organizaciones.

1.1 Entendiendo la IA: definición, evolución y transformación en Recursos Humanos[1]

En el ámbito de RR. HH., esta evolución marca un punto de inflexión: de una función tradicionalmente administrativa y descriptiva a una función predictiva y estratégica. La IA permite identificar el potencial de los empleados, anticipar necesidades de contratación, optimizar procesos de selección y desarrollar modelos personalizados de aprendizaje y desarrollo profesional. Estas capacidades están transformando los modelos de gestión del talento, generando entornos de trabajo más adaptativos, inclusivos y orientados a la mejora continua.

Para aprovechar sus oportunidades en RR. HH. es imprescindible entender qué es —y qué no es— la IA, identificar los conceptos clave que la sustentan y, sobre todo, reconocer los sesgos que pueden reproducirse a través de estas herramientas. Además, se requiere una comprensión profunda de sus limitaciones actuales, así como de las condiciones necesarias para su uso ético y responsable en la toma de decisiones sobre personas.

Comprender la IA no solo implica conocer sus capacidades, sino también sus límites actuales, sus condiciones de uso y los riesgos de una implementación acrítica. Esto es especialmente relevante en RR. HH., un ámbito en el que las decisiones afectan directamente al desarrollo profesional, la igualdad de oportunidades y el bienestar de los empleados. Garantizar un uso transparente, justo y alineado con los principios de equidad es un desafío fundamental para los responsables de gestión de personas en esta nueva era digital.

1.1.1 ¿Qué es IA y qué no es IA?

La IA puede definirse como un conjunto de sistemas capaces de realizar tareas que tradicionalmente requerían inteligencia humana (Russell & Norvig, 2021). Estas tareas incluyen el aprendizaje automático o machine learning (Mitchell, 1997), la visión por computador, el procesamiento del lenguaje natural o la toma de decisiones automatizada. En RR. HH., estas capacidades se traducen en aplicaciones muy concretas como las siguientes, entre otras:

- **Análisis predictivo para identificar,** por ejemplo, el riesgo de rotación antes de que ocurra.

[1] **Nota:** La aportación de Yolanda Gutiérrez ha sido clave en el desarrollo de este capítulo. Es socia de Consultoría en Eversheds Sutherland y miembro del Consejo Asesor del Proyecto IA+Igual.

- **Evaluación de competencias** a través de modelos de machine learning (ML) que permiten identificar patrones de desempeño y potencial de desarrollo.
- **Sistemas de recomendación de formación personalizada** que proponen itinerarios formativos en base a las habilidades del empleado y necesidades futuras de la organización.
- **Chatbots conversacionales en procesos de selección y onboarding** capaces de interactuar con candidatos para resolver dudas o gestionar solicitudes de forma automatizada y eficiente.

Un ejemplo práctico de estas aplicaciones es el sistema de People Analytics de IBM Watson, utilizado por grandes consultoras para predecir el riesgo de fuga de talento clave. Mediante el análisis de datos históricos y patrones de comportamiento, la IA anticipa con un alto grado de precisión qué empleados están en riesgo de abandonar la empresa en los próximos meses, permitiendo diseñar planes de retención específicos.

Ahora bien, no todo sistema automatizado es IA. Herramientas como los gestores de agendas, los sistemas de firma digital o el envío programado de correos electrónicos son tecnologías útiles, pero no "inteligentes" en sentido estricto. En el área de gestión de personas, un ejemplo claro son los sistemas de control horario digital que registran la entrada y salida de los empleados. Aunque optimizan el proceso de registro, no hay un aprendizaje automático ni una toma de decisiones basada en datos.

Otro ejemplo son las herramientas de seguimiento de ausencias o vacaciones: gestionan información, pero no predicen patrones, ni optimizan decisiones, ni se adaptan al comportamiento de los empleados.

Además, procesos como la firma digital de documentos laborales, aunque digitalizan y agilizan el intercambio de información, no integran capacidades de IA. La firma electrónica no aprende del comportamiento del usuario ni optimiza futuros procesos; simplemente valida una identidad y guarda un registro digital de la transacción.

La inteligencia artificial se distingue por su capacidad de aprender y adaptarse a partir de los datos (Domingos, 2015). En este sentido, un chatbot que únicamente responde con guiones predefinidos no se consideraría IA, mientras que un chatbot que aprende de cada interacción para responder mejor en el futuro, sí lo sería.

¿Tecnología objetiva o sistema con valores?

Un error común es asumir que los sistemas de IA son neutros o infalibles. En realidad, la nueva tecnología refleja las decisiones humanas que la configuran: desde los datos empleados hasta las métricas optimizadas, pasando por la lógica del modelo (Floridi & Cowls, 2019; European Commission, 2020). Estas decisiones no son únicamente técnicas, pues responden a objetivos específicos, criterios de valor implícitos y contextos institucionales concretos. Por ello, los sistemas de IA deben entenderse como construcciones sociotécnicas, es decir, tecnologías que combinan conocimiento técnico con elecciones humanas, sociales y organizativas.

Si estas decisiones no se revisan críticamente, la IA puede no solo incorporar errores, sino también reproducir y amplificar desigualdades preexistentes. Este riesgo es especialmente relevante en ámbitos como RR. HH., donde las decisiones afectan directamente a personas y trayectorias profesionales.

Entender qué es la IA en el ámbito laboral implica también comprender cómo se construyen, se

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

gobiernan y se mantienen los datos que alimentan esos sistemas. La calidad, la trazabilidad y la equidad en el tratamiento de los datos no son cuestiones técnicas secundarias: son condiciones fundamentales para garantizar una IA que respete los derechos de las personas, favorezca la inclusión social y promueva oportunidades laborales equitativas. Esta perspectiva será clave a lo largo de todo el libro, especialmente en los apartados dedicados a la gobernanza del dato, la diversidad y la normativa aplicable en el mercado laboral.

1.1.2 Antecedentes y estado del arte

La IA nace formalmente en 1956 durante la Conferencia de Dartmouth, donde científicos como John McCarthy, Marvin Minsky, Allen Newell y Herbert Simon plantearon la posibilidad de que los procesos de pensamiento humano pudieran ser descritos y replicados por máquinas (McCarthy et al., 1955; Russell & Norvig, 2021). Desde entonces, esta tecnología disruptiva ha evolucionado como un campo interdisciplinar, combinando la informática, la lógica, la estadística y la neurociencia para construir sistemas cada vez más capaces de realizar tareas de razonamiento, aprendizaje y percepción.

Evolución general de la IA

Tras una primera etapa de grandes expectativas, la IA experimentó varios "inviernos", periodos en los que el entusiasmo inicial se vio reemplazado por escepticismo debido a las limitaciones técnicas de la época. Sin embargo, el progreso en hardware, la disponibilidad masiva de datos y la mejora de los algoritmos reactivaron el interés por la IA en las últimas décadas. Especialmente desde la irrupción del aprendizaje profundo o deep learning (LeCun, Bengio & Hinton, 2015), la IA ha demostrado un rendimiento sobresaliente en tareas como el reconocimiento de imágenes, el procesamiento de lenguaje natural y la toma de decisiones automatizada.

Hoy en día, el estado del arte en IA incluye modelos de aprendizaje profundo, redes neuronales generativas, técnicas de refuerzo profundo y arquitecturas de atención como los transformadores, que sustentan sistemas tan diversos como asistentes virtuales, sistemas de recomendación, modelos de predicción médica y plataformas de generación de contenido. Esta sofisticación técnica, sin embargo, plantea interrogantes éticos y sociales que requieren enfoques de gobernanza responsables (Floridi & Cowls, 2019; European Commission, 2020) y una evaluación crítica de su impacto en las personas y las organizaciones.

RR. HH. antes de la incorporación de la IA

Antes de la incorporación de modelos de inteligencia artificial, la gestión de personas se apoyaba fundamentalmente en enfoques manuales y basados en la experiencia directa de los profesionales del área, una dinámica que ha sido ampliamente documentada en los análisis sobre transformación organizacional (Brynjolfsson & McAfee, 2014). Las decisiones respecto a procesos de reclutamiento, desarrollo y evaluación se tomaban a partir de entrevistas personales, revisiones de desempeño subjetivas y análisis relativamente simples de métricas como la rotación o la tasa de retención. Los datos existían, pero eran escasos, fragmentados y se utilizaban principalmente con fines descriptivos, no predictivos.

A partir de los años 90, con la informatización de las organizaciones, comenzaron a surgir los primeros Sistemas de información de Recursos Humanos (HRIS, por sus siglas en inglés), que priorizaban la eficiencia operativa sobre el análisis predictivo (Bersin, 2018). Estos permitieron almacenar

información de manera testructurada, facilitando la automatización de procesos administrativos como la nómina o la gestión de beneficios. Sin embargo, el análisis de esta información seguía estando orientado, principalmente, a la eficiencia operativa, sin un aprovechamiento pleno de su potencial estratégico.

Fue en las dos últimas décadas, con el auge de la cultura del dato y la adopción de tecnologías de Business Intelligence, cuando RR. HH. empezó a evolucionar hacia un enfoque más analítico (Cappelli, 2019). La disponibilidad de herramientas para medir el desempeño, la satisfacción, la formación y otros aspectos clave del talento abrió la puerta a una gestión más basada en evidencias. Este cambio de mentalidad, que colocó al dato en el centro de la toma de decisiones, sentó las bases para la adopción de soluciones de IA, que permiten no solo describir el pasado, sino también anticipar y modelar el futuro de la fuerza laboral.

Implantación de la IA en la gestión de personas

En el área de talento, la introducción de la IA no ha sido un fenómeno aislado, sino parte de un proceso de transformación digital más amplio. Inicialmente, las primeras aplicaciones se centraron en la automatización de tareas administrativas, como la criba de currículos mediante sistemas basados en reglas o el envío automatizado de invitaciones a entrevistas (Sculley et al., 2015). Una de las primeras aplicaciones destacadas fue ATS -Applicant Tracking System, por su sigla en inglés- (Upadhyay & Khandelwal, 2018). Estas plataformas, orientadas a agilizar el proceso de selección, abrieron la puerta al uso de algoritmos capaces de identificar patrones en la información de los candidatos y predecir su adecuación al puesto. Con la mejora de las capacidades de aprendizaje automático y analítica, estas soluciones evolucionaron hacia modelos que permiten anticipar necesidades de contratación, detectar riesgos de rotación o proponer trayectorias personalizadas (Jain & Singhal, 2021).

Con el avance de la analítica predictiva y el machine learning, los sistemas de gestión del talento comenzaron a evolucionar. Actualmente, la IA permite analizar grandes volúmenes de datos de empleados y candidatos, identificar patrones de desempeño, predecir el potencial de desarrollo, recomendar trayectorias formativas y medir aspectos intangibles como el clima organizacional (Bersin, 2018) o el nivel de compromiso a partir de fuentes de datos no estructurados como encuestas abiertas o registros de comunicación interna.

El estado actual de la inteligencia artificial en Recursos Humanos incluye tecnologías emergentes como la generación automática de evaluaciones de desempeño personalizadas, chatbots de asistencia al empleado basados en procesamiento de lenguaje natural y plataformas de reclutamiento que optimizan la diversidad mediante algoritmos de mitigación de sesgos. No obstante, su adopción plantea desafíos críticos: desde la necesidad de transparencia en los algoritmos hasta la gestión ética de los datos personales de los empleados (European Commission, 2020), aspectos que deben ser cuidadosamente considerados para garantizar un uso responsable y justo de estas herramientas.

1.1.3 La cultura del dato en Recursos Humanos: de los indicadores de desempeño a los modelos de IA

La transformación digital de RR. HH no puede entenderse sin la consolidación de una auténtica cultura del dato. Históricamente, los departamentos de gestión de personas han gestionado información relevante —como registros de empleados, nóminas, evaluaciones de desempeño o datos de formación—, pero el uso de estos datos se limitaba a una función operativa o descriptiva, sin aprovechar su potencial analítico (Davenport, Harris & Shapiro, 2010).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

• Evolución de la cultura del dato en RR. HH.[2]

Más allá del protagonismo que han cobrado las tecnologías digitales, lo que realmente está transformándose es la forma en la que pensamos, analizamos y decidimos dentro de las organizaciones. El paso de una lógica basada en la intuición hacia una fundamentada en la evidencia, y de un enfoque reactivo a una mirada anticipatoria, está convirtiendo a la cultura del dato en un eje estructural de los departamentos de RR. HH.

Durante décadas, el indicador por excelencia en RR. HH. fue la medición de resultados básicos: tasas de rotación, índices de absentismo, antigüedad media o costes de formación (Bassi, 2011). Estos indicadores de desempeño eran esenciales para la gestión administrativa, pero insuficientes para fundamentar decisiones estratégicas. El dato, en este contexto, era más un registro pasivo que un activo dinámico de valor.

El cambio comenzó a gestarse con la introducción de herramientas de Business Intelligence aplicadas a la gestión del talento (Levenson, 2015; Rasmussen & Ulrich, 2015). Los sistemas de HR Analytics permitieron no solo almacenar datos, sino analizarlos en búsqueda de patrones, tendencias y relaciones causales. Aparecieron nuevos indicadores, más complejos y estratégicos: tiempo de cobertura de vacantes críticas, coste de rotación de talento clave, predicción de riesgo de fuga, o tasa de éxito de programas de formación.

• Transición tecnológica y madurez analítica

El recorrido que ha seguido la cultura del dato en RR. HH. refleja una transición paulatina, no exenta de tensiones. En un primer momento, el uso de datos respondía a necesidades administrativas: gestionar contratos, registrar ausencias, calcular nóminas. Estos datos solían almacenarse en papel o, en el mejor de los casos, en hojas de Excel, sin trazabilidad ni estandarización. El verdadero punto de inflexión llegó con los HRMS (Human Resource Management Systems), que facilitaron la centralización de la información y la generación de indicadores clave de manera más fiable.

Sin embargo, la verdadera revolución ha llegado con la aplicación de técnicas de IA en los modelos de análisis de datos de Recursos Humanos. El paso de indicadores de desempeño tradicionales a modelos predictivos y prescriptivos ha cambiado el rol de los datos en la función de Personas (Marler & Boudreau, 2017; Minbaeva, 2018). Ya no se trata únicamente de observar lo que ha ocurrido, sino de anticipar comportamientos, optimizar procesos y personalizar estrategias de desarrollo y retención.

No obstante, esta transformación solo es posible si la base sobre la que se construyen los modelos de IA es robusta, coherente y confiable. Sin una infraestructura sólida de datos, la IA en RR. HH. corre el riesgo de convertirse en una promesa sin sustancia.

• Gobernanza del dato en RR. HH.

Este cambio ha venido acompañado por una creciente conciencia sobre la necesidad de establecer marcos sólidos de Gobernanza del dato. La naturaleza de la información que maneja RR. HH.

[2] **Nota:** Ana Valera, experta en People Analytics y miembro del Consejo Asesor, ha sido clave en el desarrollo de la Cultura del dato.

—profundamente sensible y vinculada a la experiencia y los derechos de las personas— exige que este departamento asuma un rol activo y propositivo en el diseño de políticas que aseguren un uso transparente, equitativo y explicable de los algoritmos.

La adopción de esta nueva cultura del dato no es homogénea entre las organizaciones. Muchas empresas aún se encuentran en fases incipientes, centradas en el fortalecimiento de sus sistemas de captura y gestión de información (Insights for Professionals, 2021). Otras, más avanzadas están explorando la construcción de modelos de predicción de rotación, mapeo de potencial interno o recomendación de itinerarios de carrera personalizados.

La transición hacia una cultura analítica sólida requiere más que herramientas tecnológicas. Exige un cambio en las competencias de los profesionales de RRHH, en la mentalidad directiva y en la gobernanza de los datos. Sin esta evolución cultural, las inversiones en IA aplicadas a la gestión de personas corren el riesgo de quedar infrautilizadas (Bersin, 2018; IBM Smarter Workforce Institute, 2017) o, peor aún, de generar resultados sesgados y contrarios a los principios de equidad y desarrollo humano que deben guiar la función de RR. HH.

Este cambio de enfoque, del uso del dato como instrumento a su integración en modelos inteligentes, prepara el camino hacia una transformación más profunda del área de RRHH, donde la IA no solo informa las decisiones, sino que las redefine.

1.1.4 La IA como factor transformador en RR. HH.

La IA representa un cambio estructural en la manera en que se concibe y se gestiona el área de RR. HH. (Tambe, Cappelli & Yakubovich, 2019). No se trata solo de una herramienta que mejora la eficiencia, sino de un agente que redefine el papel estratégico de la función de personas dentro de las organizaciones. La IA transforma los procesos, las competencias requeridas, la toma de decisiones y, en última instancia, la cultura organizativa.

Tradicionalmente, RR. HH. ha sido una función de soporte, centrada en la administración del personal, el cumplimiento normativo y la intermediación en conflictos laborales. Con la llegada de la IA, esta función puede evolucionar hacia un rol más estratégico: analizando datos para anticipar necesidades de talento, diseñando programas personalizados de desarrollo profesional o promoviendo una cultura basada en evidencia y mejora continua (Minbaeva, 2020).

La IA permite modelizar comportamientos, predecir resultados y automatizar procesos complejos, lo que supone un salto cualitativo respecto a los enfoques anteriores. Esto impacta en áreas como la selección, el onboarding, el aprendizaje continuo, la gestión del desempeño o la planificación de la sucesión (Daugherty & Wilson, 2018; Bersin, 2019) entre otros. Pero su efecto más profundo no está solo en la mejora técnica de estos procesos, sino en la posibilidad de reconfigurar la lógica misma de cómo se conciben y gestionan. Además, ofrece herramientas para detectar oportunidades de mejora en aspectos clave como la diversidad, la equidad y la inclusión, minimizando los sesgos que históricamente han condicionado procesos de selección y evaluación (Buolamwini & Gebru, 2018; Noble, 2018).

El papel transformador de la IA exige también una transformación de los perfiles profesionales dentro del área del talento. Las competencias técnicas, analíticas y estratégicas adquieren un nuevo peso, y se requiere una combinación de sensibilidad organizativa y alfabetización tecnológica (OECD, 2021;

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

World Economic Forum, 2023) para liderar esta transición.

Además, la integración de la IA desafía las estructuras organizativas tradicionales. Requiere nuevas formas de colaboración interdisciplinar, gobernanza compartida y diseño ético de procesos (IBM, 2021; Ulrich & Dulebohn, 2015). Esta dimensión ética ha sido ampliamente abordada por autoras como Eubanks (2018), que alertan sobre el riesgo de automatizar desigualdades si no se garantiza una supervisión humana y una gobernanza transparente. El área de personas ya no puede limitarse a observar cómo la tecnología transforma la organización, debe posicionarse como actor clave en la gestión del cambio.

Actualmente, la aplicación de IA en Recursos Humanos se encuentra en una fase de consolidación, con una creciente demanda de soluciones más responsables, auditables y centradas en las personas (Dhamija, 2021). Esta evolución marca el paso de un enfoque meramente tecnológico a uno más estratégico y humanista, en el que la IA actúa como un apoyo para la creación de entornos de trabajo más inclusivos, sostenibles y orientados al bienestar de las personas trabajadoras (Guenole, Ferrar & Feinzig, 2017; Huang & Rust, 2021).

Por todo ello, los departamentos de RR. HH. no son únicamente usuarios de estas herramientas, sino garantes de que su implementación contribuya a construir entornos de trabajo más justos, diversos y sostenibles (Mehrabi et al., 2021; Eubanks, 2018). Este papel transformador que la IA puede desempeñar en RRHH ha guiado también la estructura del presente Libro Blanco. Para ello, se han definido tres ejes estratégicos que articulan la acción pública y permiten enmarcar los retos y oportunidades que emergen con el uso de la IA en el ámbito laboral. Estos ejes —promoción de la autonomía e inclusión social, transformación digital e inserción sociolaboral— se han interpretado como principios estructurales que orientan tanto la selección de contenidos como la propuesta de buenas prácticas, certificaciones y recomendaciones.

Cada eje estratégico se traduce en un conjunto de principios aplicables a la función de RR. HH.: desde garantizar la igualdad de acceso al empleo y la equidad en la selección de personal, hasta reducir la brecha digital, promover los derechos digitales y acompañar a colectivos vulnerables en procesos de inserción laboral mediante sistemas libres de sesgos. La integración de estos ejes no es un añadido técnico, sino un enfoque transversal que subraya que la transformación que plantea la IA no será completa si no es, al mismo tiempo, social, ética e inclusiva.

La IA en RR. HH. no es, por tanto, una fase técnica dentro de una secuencia de digitalización. Es un punto de inflexión que redefine la naturaleza misma de la función. Este potencial transformador plantea oportunidades —pero también responsabilidades— para que RR. HH. no solo acompañe, sino lidere el proceso de construcción de organizaciones más ágiles, inclusivas y centradas en las personas (McKinsey & Company, 2022; CIPD, 2020).

1.2 Beneficios y retos en la gestión de personas

La adopción de la inteligencia artificial en el ámbito de los Recursos Humanos representa uno de los vectores más significativos de transformación organizativa en la actualidad. Desde la mejora operativa hasta la redefinición de modelos de liderazgo, la inteligencia artificial está reconfigurando las prácticas, roles y valores que sostienen la gestión del talento. Sin embargo, sus efectos no son homogéneos ni unidimensionales: junto a los beneficios emergen tensiones estructurales, dilemas éticos y nuevas exigencias normativas.

Este apartado examina, desde una perspectiva progresiva, los principales beneficios, tendencias globales, necesidades formativas, riesgos éticos y marcos de cumplimiento que condicionan el uso responsable de la IA en la gestión de personas. Algunos subapartados presentan un desarrollo más extenso, en la medida en que abordan dimensiones estratégicas especialmente sensibles o estructurales, como la equidad, los sesgos o la evolución de competencias clave. Todos ellos comparten un mismo objetivo: ofrecer una lectura crítica y orientada a la acción sobre los desafíos reales que plantea la automatización en los entornos laborales.

La IA aplicada a RR. HH. ha generado una transformación significativa en múltiples procesos: automatización del cribado de candidatos, personalización de itinerarios formativos, optimización de la experiencia del empleado y monitorización del clima organizativo, entre otros. Estas herramientas no solo incrementan la eficiencia operativa, sino que también amplían la capacidad estratégica del área de personas. Lejos de reemplazar el juicio humano, la IA puede actuar como un soporte analítico y adaptativo que potencia la toma de decisiones, siempre que se integre bajo principios de equidad, transparencia y supervisión responsable.

1.2.1 Impacto de la IA en la gestión de personas

La incorporación de la IA en la gestión de personas no es un fenómeno puntual, sino la expresión de un cambio más profundo en el modo en que se entienden, gestionan y optimizan los procesos de RR. HH. La IA actúa como catalizador de una transición que va más allá de la digitalización operativa: permite introducir capacidades analíticas, predictivas y adaptativas que transforman la función de RR. HH. desde sus cimientos (Minbaeva, 2020).

En una primera etapa, el impacto de la IA se ha manifestado principalmente en la automatización de tareas repetitivas: la criba de currículos mediante algoritmos de aprendizaje supervisado, el análisis sintáctico y semántico de cartas de motivación, o el uso de chatbots para responder preguntas frecuentes en procesos de onboarding (Upadhyay & Khandelwal, 2018). Estas aplicaciones han permitido reducir cargas administrativas y aumentar la eficiencia operativa, pero su alcance ha sido, hasta ahora, fundamentalmente instrumental.

Sin embargo, el potencial real de la IA en RR. HH. se despliega cuando se combina con estrategias de analítica avanzada del talento. A través del uso de grandes volúmenes de datos estructurados y no estructurados —encuestas internas, evaluaciones de desempeño, interacción en herramientas colaborativas— es posible construir modelos que anticipen riesgos de rotación, identifiquen trayectorias de alto rendimiento o recomienden itinerarios formativos personalizados (Bersin, 2018). Este enfoque, conocido como people analytics predictivo, permite pasar de una lógica de observación del pasado a una capacidad real de anticipación estratégica (Jain & Singhal, 2021).

Ahora bien, no se trata solo de analizar datos. El uso de IA en la gestión de personas obliga a repensar el equilibrio entre eficiencia tecnológica y juicio humano. No debe concebirse como un sustituto de la toma de decisiones humanas, sino como un soporte que permite ampliar la información disponible, aportar mayor claridad y mejorar la objetividad en los procesos. Este enfoque exige garantizar que los modelos utilizados respeten principios éticos y derechos laborales, especialmente cuando las decisiones afectan directamente al ciclo de vida profesional de las personas (Eubanks, 2018; European Commission, 2020).

La transformación que introduce la IA también tiene una dimensión cultural. El principal obstáculo a su implantación no es tanto de tipo técnico como organizacional. Muchas empresas aún no han

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

desarrollado una cultura del dato sólida que permita aprovechar el verdadero potencial de estas herramientas.

Persisten barreras como la falta de competencias analíticas en los equipos de RR. HH., la desconfianza hacia decisiones automatizadas o el desconocimiento generalizado de las capacidades reales de la IA más allá de los discursos aspiracionales (CIPD, 2021).

No obstante, el escenario está cambiando. A medida que los modelos se refinan y las plataformas se integran mejor en los sistemas de gestión de talento, la IA comienza a generar valor en áreas como la personalización de la experiencia del empleado, la promoción de la diversidad mediante algoritmos de mitigación de sesgos o la planificación dinámica de la fuerza laboral en función de proyecciones de negocio (World Economic Forum, 2023).

En este sentido, la incorporación de herramientas digitales accesibles, como asistentes virtuales o entornos formativos inteligentes, puede desempeñar un papel relevante en la reducción de la brecha digital, especialmente entre jóvenes o personas en situación de vulnerabilidad.

Diseñadas adecuadamente, estas soluciones no solo agilizan procesos, sino que también promueven la inclusión digital y el ejercicio pleno de los derechos en entornos laborales altamente tecnologizados.

La clave, en última instancia, no reside en la tecnología en sí misma, sino en cómo se integra, se gobierna y se pone al servicio de una visión estratégica de la gestión de personas. Como se argumentará en los apartados dedicados a la ética y a la regulación, la implementación responsable de la inteligencia artificial en el ámbito laboral y de los recursos humanos requiere transparencia, trazabilidad y supervisión humana permanente.

Solo así será posible aprovechar su capacidad transformadora sin renunciar a los principios de equidad, inclusión y respeto que deben guiar toda política de personas (Brynjolfsson & McAfee, 2014; McKinsey & Company, 2022).

1.2.2 Tendencias en Europa y en el mundo

La adopción de la IA en los procesos de gestión de personas está avanzando de forma desigual entre regiones, sectores y culturas organizativas. Aunque la tecnología es global, sus formas de implementación, los marcos regulatorios y los enfoques éticos que la acompañan responden a realidades profundamente contextualizadas. Por ello, observar las tendencias actuales en Europa y en el mundo permite comprender no solo el estado del arte, sino también las tensiones y dilemas que acompañan su despliegue en el ámbito laboral (World Economic Forum, 2023).

En Europa, la integración de la IA en RR. HH. ha estado marcada por una fuerte preocupación por los derechos fundamentales, la protección de datos y la equidad. El GDPR y el recién aprobado RIA establecen un marco jurídico que clasifica las aplicaciones de IA en RR. HH. como de alto riesgo, exigiendo transparencia, trazabilidad y supervisión humana (European Commission, 2023).

Este enfoque ha favorecido una implementación más cautelosa, pero también más alineada con los principios de una IA confiable promovida por organismos como la OCDE y la UNESCO (OECD, 2021; UNESCO, 2021).

Tabla 1: *Áreas clave de impacto de la IA en la gestión de personas*

Área funcional	Aplicaciones IA	Valor añadido	Riesgos clave
Selección y reclutamiento	Criba automática CV, matching algorítmico, análisis entrevistas	Eficiencia y reducción de sesgos humanos	Reproducción de sesgos históricos
Formación y desarrollo	Rutas formativas personalizadas Análisis de aprendizaje	Aprendizaje adaptativo Mayor compromiso	Pérdida de autonomía o perfilado invasivo
Desempeño y carrera	Evaluaciones automatizadas Análisis predictivo potencial	Más objetivas Planificación estratégica	Falta de explicabilidad
Clima y bienestar	Análisis de sentimiento Detección de burnout	Prevención temprana Programas proactivos	Uso inapropiado de datos sensibles
Planificación de talento	Modelos de rotación Identificación de líderes	Optimización recursos Retención	Dependencia de modelos inestables

Fuente: elaboración propia.

Tabla 2: *Comparativa de uso de IA en RR. HH. según región*

Región	Lógica adopción	Foco principal	Nivel regulación	Tipo implementación	Riesgos observados
Europa	Precaución estratégica	Derechos Equidad Confianza	Alta RIA, GDPR	Gradual, ético, regulado	Desigualdad por madurez digital entre sectores
EE. UU.	Innovación competitiva	Escalabilidad Productividad	Baja (autorregulación)	Rápida, liderada por grandes tecnológicas	Falta transparencia Sesgos no auditados
Asia	Eficiencia y control	Automatización masiva Rendimiento	Variable (modelos estatales o empresariales)	Masiva en sectores clave, con poca supervisión externa	Riesgos en vigilancia laboral y perfilado

Fuente: elaboración propia.

A nivel empresarial, son los sectores con mayor madurez digital —como la tecnología, la consultoría o la industria avanzada— los que lideran la adopción de IA en el ámbito laboral. En cambio, en las pymes y el sector público, la implementación sigue siendo más fragmentaria, con desarrollos aún incipientes. Esta situación se ve reflejada en los índices internacionales de adopción de IA, que muestran una clara correlación entre infraestructura digital y despliegue tecnológico en RR. HH. (McKinsey & Company, 2022).

En contraposición, en Estados Unidos y gran parte de Asia, la lógica predominante ha sido la de la eficiencia, la escalabilidad y la ventaja competitiva. Las grandes tecnológicas han liderado la automatización del ciclo de vida del empleado, incluyendo procesos como la selección, la monitorización del desempeño o el diseño de rutas de formación algorítmicas. Esta aceleración ha venido acompañada de preocupaciones sobre el uso de datos sensibles y la falta de explicabilidad de los algoritmos utilizados (AI Now Institute, 2020). A diferencia del modelo europeo, los marcos éticos en estos entornos provienen en gran medida de autorregulación empresarial o consorcios sectoriales, más que de una normativa pública fuerte (Dutton, 2018).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

A nivel global, se observa una tendencia convergente: la consolidación de plataformas híbridas que combinan automatización, aprendizaje automático y supervisión humana. Estas soluciones buscan equilibrar eficiencia y sensibilidad organizativa, especialmente en tareas como el reclutamiento, la formación personalizada o la predicción de fuga de talento (Gartner, 2023). Además, comienza a despegar el uso de IA generativa en procesos de asistencia al empleado, documentación interna o generación de contenidos formativos, lo cual plantea nuevas preguntas sobre el papel de estas herramientas en la arquitectura de conocimiento de las organizaciones (Harvard Business Review, 2021).

Este despliegue tecnológico se expresa también en tendencias funcionales concretas que están moldeando la práctica diaria de los RR. HH. La integración de modelos generativos en la comunicación interna, el uso de analítica predictiva para anticipar rotación o rendimiento, y la adopción de asistentes virtuales para automatizar procesos administrativos son ya realidades en muchas organizaciones (Harvard Business Review, 2023; OECD, 2023). Asimismo, se afianzan aplicaciones orientadas a fomentar la diversidad y la inclusión, mediante herramientas que identifican sesgos en tiempo real o recomiendan intervenciones correctivas (Tambe, Cappelli & Yakubovich, 2023). Estas tendencias también abren posibilidades para el desarrollo de formatos accesibles y personalizados de aprendizaje y orientación laboral. Tecnologías como la gamificación, la IA conversacional o los entornos inmersivos pueden adaptarse a las necesidades de colectivos vulnerables y jóvenes en riesgo de exclusión, siempre que se diseñen bajo principios de inclusión digital y equidad tecnológica.

Estas innovaciones, si bien se extienden a escala global, adoptan formas distintas según el contexto normativo, la madurez digital o la cultura organizativa. En paralelo, iniciativas multilaterales como el Global Partnership on AI, los Principios de la OCDE o la Carta Europea de Derechos Digitales evidencian un esfuerzo por construir un marco ético compartido a escala internacional (GPAI, 2022; OECD, 2021). Estas propuestas coinciden en la necesidad de garantizar que el despliegue de IA se guíe por valores democráticos, la protección de los derechos humanos y el objetivo de una transición digital justa e inclusiva.

En conjunto, el panorama internacional refleja una tensión entre aceleración y cautela, entre eficiencia y equidad. Comprender esta pluralidad de enfoques no solo permite identificar buenas prácticas, sino también anticipar los dilemas que enfrentará cualquier organización que aspire a integrar la inteligencia artificial de forma estratégica, ética y sostenible en la gestión de personas.

1.2.3 Necesidad de reskilling y upskilling

La transformación impulsada por la IA no se limita a procesos, herramientas o decisiones: afecta al núcleo mismo de las competencias humanas necesarias para operar y liderar en el nuevo contexto laboral. En el ámbito de RR. HH., esta transformación es doble: por un lado, implica adaptar las capacidades del propio departamento para trabajar con sistemas inteligentes; por otro, exige anticipar, diseñar y facilitar estrategias de aprendizaje continuo para el conjunto de la organización (OECD, 2021).

La distinción entre reskilling (reciclaje profesional) y upskilling (actualización de capacidades) no es meramente técnica: refleja dos formas complementarias de responder a una realidad en mutación. Mientras el reciclaje aborda la necesidad de adquirir competencias totalmente nuevas —por ejemplo, cuando un perfil operativo debe orientarse hacia funciones digitales—, la actualización se centra en

profundizar o ampliar conocimientos existentes —como la capacitación de líderes en el uso ético de sistemas de IA en la toma de decisiones organizativas (Harvard Business Review, 2020).

En este contexto, RR. HH. no puede limitarse a ejecutar programas formativos reactivos. La función debe adoptar un rol estratégico, transversal y anticipador, capaz de mapear el talento disponible, detectar brechas críticas de competencias y proponer itinerarios personalizados que integren tecnologías emergentes, metodologías activas y principios de inclusión. Esta misión exige superar la lógica tradicional de la formación como evento aislado y avanzar hacia un ecosistema de aprendizaje permanente, basado en datos, diseñado con las áreas de negocio y alineado con los objetivos de transformación digital (IBM Institute for Business Value, 2022).

Los estudios internacionales coinciden en señalar que las habilidades técnicas vinculadas a la IA —como la alfabetización digital, la analítica de datos o el pensamiento computacional— son cada vez más relevantes, pero no suficientes. Tan importante como saber manejar una herramienta es entender su impacto organizacional, social y ético. Por ello, junto a las competencias técnicas, emergen con fuerza habilidades transversales como la adaptabilidad, la inteligencia emocional, el juicio crítico y la capacidad de trabajar en entornos híbridos y colaborativos (World Economic Forum, 2023).

Identificar qué habilidades serán realmente críticas implica reconocer que el valor diferencial no radica solo en la competencia técnica. Las organizaciones deberán fomentar una combinación equilibrada entre capacidades digitales, pensamiento analítico y habilidades sociales complejas, como la empatía, la comunicación o el juicio ético (Harvard Business Review, 2021).

Este cambio de paradigma afecta directamente al perfil de los profesionales de RR. HH. Ya no basta con dominar las políticas laborales o los procesos administrativos. Se requiere una alfabetización tecnológica aplicada, un conocimiento básico del funcionamiento de los algoritmos y una sensibilidad especial para traducir la lógica de la IA a decisiones que respeten la equidad y la diversidad dentro de las organizaciones. Esto implica, entre otras cosas, poder formular las preguntas adecuadas a los equipos técnicos, interpretar salidas algorítmicas y detectar posibles sesgos en la aplicación de modelos predictivos (CIPD, 2021).

En este sentido, la utilización de marcos europeos como ESCO o DigComp puede resultar especialmente valiosa. El primero —European Skills, Competences, Qualifications and Occupations— permite definir perfiles profesionales en términos de competencias observables y alineadas con el mercado laboral europeo (European Commission, 2023a). El segundo —Digital Competence Framework— establece una taxonomía clara y graduada de competencias digitales, agrupadas en cinco áreas: alfabetización digital, comunicación, creación de contenido, seguridad y resolución de problemas (European Commission, 2023b). Ambos instrumentos ofrecen un lenguaje común para diseñar itinerarios formativos, mapear capacidades internas y evaluar progresos de forma sistemática y comparable a nivel internacional (Cedefop, 2021).

Asimismo, la implantación de procesos de IA en el ciclo de vida del empleado exige una revisión de los modelos tradicionales de evaluación y promoción. Si el aprendizaje continuo se convierte en una de las áreas de la empleabilidad, entonces los sistemas de reconocimiento, recompensa y progreso profesional deben estar alineados con ese principio. Esto incluye valorar no solo los resultados, sino también la capacidad de aprendizaje, la disposición al cambio y la contribución al desarrollo colectivo del conocimiento interno (OECD, 2022).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Tabla 3: *Áreas clave del marco Competencias digitales en el entorno laboral*

Competencia digital	Aplicación RR. HH.	Nivel deseable según rol
Alfabetización digital	Buscar, evaluar y gestionar información laboral y normativa	Todos los empleados
Comunicación y colaboración	Uso de plataformas colaborativas, comunicación remota, gestión de equipos híbridos	Mandos intermedios, técnicos y líderes
Creación de contenido digital	Redacción de informes interactivos, generación de contenido formativo con IA	Técnicos en formación, comunicación, innovación
Seguridad	Protección de datos personales, ciber higiene, cumplimiento GDPR	Toda la organización -foco en RR. HH.| IT-
Resolución de problemas	Identificación de oportunidades tecnológicas, uso crítico de IA, adaptación a nuevos sistemas	Equipos de transformación, perfiles estratégicos

Fuente: elaboración propia.

Además del enfoque estratégico y normativo, resulta útil considerar ejemplos de aplicación práctica del reskilling y el upskilling dentro de las funciones clave de los RR. HH. A continuación, se recoge una propuesta metodológica que permite visualizar cómo estas estrategias pueden materializarse de forma concreta en los distintos procesos del ciclo de vida del talento.

Aplicaciones prácticas del reskilling y el upskilling en RR. HH.

- **Reclutamiento y selección:** adaptación de perfiles a nuevos requerimientos digitales; uso de IA para predecir potencial de aprendizaje más allá de la experiencia previa.
- **Onboarding:** diseño de itinerarios formativos personalizados desde el ingreso, con acompañamiento inteligente basado en IA conversacional.
- **Formación y desarrollo:** identificación de brechas de competencias mediante analítica de datos; implementación de programas continuos con micro learning y gamificación.
- **Evaluación del desempeño:** incorporación de nuevas métricas vinculadas a adaptabilidad, aprendizaje continuo y colaboración digital.
- **Retención del talento:** desarrollo de rutas de carrera personalizadas; uso de IA para detectar señales de desmotivación o desalineación con los nuevos entornos.

Esta metodología[2] también contempla la emergencia de nuevos roles en RR. HH. como el Talent Intelligence Specialist, el Digital Learning Designer o el Human-AI Collaboration Facilitator, que actúan como catalizadores del cambio.

Este tipo de enfoques operativos permite traducir la estrategia en acciones concretas, facilitando así una cultura organizativa orientada al aprendizaje, al desarrollo inclusivo y a la adaptación sostenible ante la disrupción tecnológica.

Por último, conviene subrayar que la estrategia de reskilling y upskilling no puede entenderse como

[3] **Nota**: Esta propuesta metodológica ha sido elaborada por Olga Herrero, experta en Formación y miembro del Consejo Asesor del Proyecto IA+Igual.

una responsabilidad individualizada del trabajador. Las organizaciones deben generar las condiciones estructurales, culturales y materiales que faciliten el aprendizaje, reduzcan las barreras de acceso y promuevan la participación de colectivos tradicionalmente excluidos de los programas de desarrollo (Unesco, 2021). En este sentido, la IA puede ser aliada o riesgo: puede facilitar la personalización y el acceso a micro – aprendizajes adaptativos, pero también puede agravar las brechas si su despliegue no se hace desde una lógica inclusiva y proactiva. Para que las estrategias de reskilling y upskilling sean sostenibles y equitativas, es imprescindible ir más allá del marco organizativo. Muchas personas trabajadoras se enfrentan a barreras de acceso a la formación, ya sea por limitaciones económicas, de tiempo o conectividad. En este sentido, las políticas públicas, los sistemas educativos y las alianzas entre empresas e instituciones juegan un papel determinante en la creación de ecosistemas formativos accesibles, inclusivos y eficaces (ILO, 2021; McKinsey & Company, 2021).

Frente a una transformación tecnológica de esta magnitud, el verdadero diferencial competitivo no será la posesión de la herramienta más avanzada, sino la capacidad de aprender, desaprender y volver a aprender de forma continua. En este escenario, la función de RR. HH. no solo acompaña el cambio: lo hace posible.

Estas estrategias de formación deben contemplar también las necesidades específicas de personas en situación de vulnerabilidad, como mujeres víctimas de violencia de género, jóvenes tutelados, migrantes o personas con discapacidad. La utilización de metodologías accesibles y formatos innovadores —como el aprendizaje por módulos, la gamificación o la IA generativa— puede facilitar la adquisición de competencias en entornos seguros, flexibles y adaptados, reforzando así los principios de equidad, autonomía y empleabilidad sostenida.

Sin embargo, acompañar esta transformación con estrategias formativas no es suficiente. A medida que se extiende el uso de IA en la gestión de personas, emergen nuevos desafíos éticos que deben ser anticipados y gestionados con responsabilidad.

1.2.4 Introducción a los retos y consideraciones éticas

El uso de la IA en el ámbito de los RR. HH. abre posibilidades que hace apenas una década parecían inalcanzables: procesos de selección más ágiles, trayectorias de aprendizaje personalizadas, análisis predictivo del desempeño, optimización del talento y diseño dinámico de políticas de bienestar. Sin embargo, junto a estos avances, emergen con fuerza nuevas preguntas éticas, sociales y organizativas que exigen ser tratadas con rigor y proactividad (High-Level Expert Group on AI, 2019).

A diferencia de otras tecnologías, la IA no es neutral ni inocua. Su aplicación en RR. HH. incide directamente en decisiones que afectan la dignidad, el desarrollo y las oportunidades de las personas. Cuando un algoritmo filtra candidaturas, recomienda un ascenso o identifica un posible riesgo de fuga, no solo actúa sobre datos: actúa sobre trayectorias de vida (Eubanks, 2018). Esta centralidad del factor humano convierte el despliegue de IA en un espacio de tensión entre eficiencia operativa y responsabilidad institucional.

Los riesgos éticos no son abstractos. Están documentados en estudios y casos reales: desde sesgos algorítmicos que perpetúan desigualdades de género o raza, hasta sistemas opacos que dificultan la rendición de cuentas o tecnologías de monitorización que erosionan la confianza y la autonomía de los trabajadores (Jobin et al., 2019; AI Now Institute, 2019). En muchos casos, los modelos utilizados por las organizaciones operan como verdaderas “cajas negras”, lo que significa que ni siquiera los propios

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

responsables de RR. HH. comprenden del todo cómo se generan las recomendaciones o decisiones (Whittlestone et al., 2019). Esta opacidad tecnológica impide una trazabilidad efectiva y limita la capacidad de intervención cuando se producen resultados injustos o inesperados. La mayoría de estos problemas no surgen de una mala intención explícita, sino de la falta de evaluación crítica en el diseño, la implementación o el uso de los sistemas (Mittelstadt et al., 2016).

Tabla 4: *Tipología de retos éticos emergentes en la aplicación de IA a RR. HH.*

Dimensión ética afectada	Ejemplo aplicación IA	Riesgo asociado	Necesidad intervención
Justicia y equidad	Algoritmos selección de personal	Reproducir sesgos históricos	Evaluación datos de entrenamiento | resultados
Transparencia	Modelos de evaluación de desempeño opacos	No explicable ante decisiones adversas	Uso de IA explicable
Autonomía y dignidad	Sistemas monitorización del comportamiento	Percepción de vigilancia Erosión de la confianza	Definir límites éticos uso IA
Responsabilidad	Decisiones automatizadas sin supervisión clara	Difícil asignación de responsabilidad errores	Establecer canales de gobernanza | trazabilidad
Privacidad	Análisis comunicaciones internas	Tratamiento indebido de datos sensibles	Aplicar el principio de minimizar datos

Fuente: elaboración propia.

La aparición de estos dilemas ha generado un consenso creciente: el avance tecnológico no puede ir por delante del juicio ético. La regulación —como se abordará más adelante— es una respuesta necesaria, pero no suficiente. Las organizaciones deben dotarse de marcos de gobernanza ética, formación específica, protocolos de revisión y canales de participación que aseguren que el uso de IA en RR. HH. esté al servicio de las personas y no a su costa (Floridi & Cowls, 2019; Unesco, 2021).

En este contexto, los retos éticos no son obstáculos al progreso, sino condiciones del mismo. Comprenderlos, debatirlos y anticiparlos es la base sobre la cual se podrá construir un uso de la IA que fortalezca la equidad, respete la diversidad y promueva una cultura organizativa más justa y sostenible (OECD, 2021; Binns, 2018). Como señalan diversos estudios, existe una brecha persistente entre los marcos éticos formulados a nivel institucional y su implementación práctica en entornos reales (Jobin, Ienca & Vayena, 2019). Reducir esa distancia requerirá no solo normas, sino también metodologías claras, capacitación transversal y una cultura organizativa comprometida con el uso justo y responsable de la tecnología.

A continuación, se desarrollarán de forma específica las implicaciones éticas y legales de la IA en los distintos procesos de RR. HH., a través de los apartados dedicados a sesgos, equidad, transparencia, privacidad, regulación y estrategias de cumplimiento (European Commission, 2020; Zarsky, 2016).

1.2.5 Sesgos inconscientes y sesgos en los algoritmos

A medida que se intensifica el uso de IA en la gestión del talento, uno de los retos más complejos que afrontan las organizaciones es la identificación y mitigación de sesgos —tanto humanos como algorítmicos— que comprometen la equidad en los procesos decisivos.

a) Sesgos inconscientes en la toma de decisiones

En el contexto de la IA aplicada a la gestión del talento, uno de los aspectos más relevantes —y a menudo menos visibles— es el papel que juegan los sesgos en la toma de decisiones. Estos sesgos pueden originarse en el juicio humano o incorporarse a los sistemas automatizados. Antes de abordar cómo se manifiestan en los algoritmos, es fundamental comprender cómo operan en las personas.

Tabla 5: *Tipos de sesgos inconscientes*

Tipo de sesgo	Descripción	Ejemplo	Referencias clave
Afinidad	Tendencia a favorecer a personas con las que compartimos antecedentes, intereses o experiencias	Preferir a un candidato que fue a la misma universidad o que comparte un hobby	Greenwald & Banaji (1995) Rivera (2012)
Efecto halo	Una característica positiva influye en la percepción global de una persona	Pensar que alguien es inteligente solo porque habla con seguridad	Thorndike (1920) Nisbett & Wilson (1977)
Efecto cuerno	Una característica negativa domina la evaluación general.	Desestimar a un candidato por una falta menor en su CV	Nisbett & Wilson (1977)
Confirmación	Tendencia a buscar o interpretar información que confirme nuestras creencias previas	Solo prestar atención a los datos que refuerzan una impresión inicial.	Nickerson (1998) Kahneman (2011)
Género	Asociar inconscientemente ciertos roles o competencias a un género específico	Asumir que los hombres son más aptos para trabajos técnicos	Moss-Racusin et al. (2012) Heilman (2001)
Edad	Juicios basados en estereotipos asociados a la edad	Considerar que una persona mayor no es suficientemente adaptable	Posthuma & Campion (2009)
Racial o étnico	Inferencias o expectativas basadas en el origen racial o étnico.	Inferir menor cualificación por el nombre en un CV	Bertrand & Mullainathan (2004)
Estatus socioeconómico	Juicios condicionados por la percepción del nivel económico o social	Asumir que alguien con vestimenta sencilla es menos profesional	Kraus & Keltner (2009)
Belleza	Atribución de rasgos positivos a personas atractivas físicamente	Considerar que alguien atractivo es más competente	Dion, Berscheid & Walster (1972)
Contraste	Evaluar a una persona en función de otra previamente evaluada, en lugar de con base en criterios objetivos	Un candidato promedio parece excelente si el anterior era muy débil	Kahneman & Tversky (1974)

Fuente: elaboración propia.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Los procesos de selección, evaluación y promoción del talento están profundamente influenciados por mecanismos automáticos de pensamiento que operan sin que seamos plenamente conscientes de ellos. Estos sesgos inconscientes corresponden a mecanismos automáticos de pensamiento que operan en los procesos de selección, evaluación y promoción del talento, sin una consciencia explícita por parte de la persona que los manifiesta. Sin embargo, en los últimos años, muchas organizaciones han comenzado a introducir metodologías específicas para minimizar estos sesgos, mediante técnicas de evaluación ciega, procesos automatizados de revisión de candidaturas y programas de formación en diversidad e inclusión.

Estos sesgos se basan en estereotipos culturales, experiencias previas o normas sociales interiorizadas, y afectan la forma en que percibimos y evaluamos a otras personas (Steele, 2010; Noble, 2018). En contextos laborales, pueden influir negativamente en la diversidad, limitar el desarrollo del talento y reproducir desigualdades existentes, incluso en organizaciones comprometidas con la equidad (Eubanks, 2018; Bielby, 2000; Williams, 2014).

Reconocer la existencia de sesgos inconscientes es el primer paso para poder mitigarlos. En las últimas décadas, numerosos estudios los han identificado y clasificado, especialmente en el ámbito de Recursos Humanos, ofreciendo herramientas para su análisis y abordaje. La anterior tabla resume los más frecuentes en contextos laborales, junto con una breve descripción, ejemplos prácticos y referencias clave que han contribuido a su estudio.

Autores clave en el estudio de los sesgos inconscientes

A lo largo de las últimas décadas, diversas figuras académicas han sido fundamentales para comprender cómo operan los sesgos inconscientes, especialmente en entornos laborales y organizacionales. Sus aportaciones han permitido conceptualizar estos mecanismos, medir su impacto y proponer estrategias de mitigación.

- **Daniel Kahneman,** *Thinking, Fast and Slow* (2011): Premio Nobel de Economía, analizó cómo los sistemas automáticos de pensamiento conducen a errores sistemáticos de juicio.

- **Mahzarin R. Banaji y Anthony G. Greenwald,** *Blindspot: Hidden Biases of Good People* (2013): estudiaron cómo incluso las personas con buenas intenciones pueden mantener sesgos implícitos sin ser conscientes de ello.

- **Claude Steele,** *Whistling Vivaldi* (2010): introdujo el concepto de amenaza del estereotipo (stereotype threat), mostrando cómo las expectativas sociales afectan el rendimiento y la autopercepción.

- **William T. Bielby:** Investigó cómo las estructuras y prácticas organizacionales pueden reforzar sesgos de género y raza en la toma de decisiones.

- **Joan C. Williams:** Estudió los sesgos que afectan la percepción de competencia en mujeres y otros grupos subrepresentados, así como sus efectos en la trayectoria profesional.

Estas referencias ofrecen una base sólida para comprender la dimensión psicológica y social de los sesgos, así como su impacto en la equidad dentro de las organizaciones.

¿Qué relación tienen con la IA?

La importancia de los sesgos inconscientes en este contexto es doble. Por un lado, siguen presentes incluso cuando se implementan sistemas automatizados. Por otro, y más preocupante aún, pueden trasladarse a los algoritmos a través de los datos históricos con los que se entrenan. Es decir, la IA puede aprender de nuestras decisiones sesgadas del pasado y reproducir esos patrones a gran escala si no se supervisa adecuadamente.

Además, incluso cuando los datos parecen estar libres de sesgos evidentes, los algoritmos pueden amplificar patrones no deseados o crear nuevas desigualdades a través de mecanismos de retroalimentación. Investigaciones recientes han demostrado que los modelos de IA pueden intensificar sesgos a partir de interacciones humanas o bucles de datos, lo que refuerza la necesidad de una supervisión constante y exhaustiva del modelo (Taori & Hashimoto, 2022; Nature Human Behaviour, 2021).

Consideraciones finales

Comprender la existencia y el funcionamiento de los sesgos inconscientes es fundamental para construir procesos de gestión del talento más justos, inclusivos y eficaces. Estos mecanismos, aunque invisibles a primera vista, tienen consecuencias tangibles: pueden restringir la diversidad, afectar a la equidad en la evaluación del desempeño y limitar el acceso a oportunidades profesionales (Kahneman, 2011; Moss-Racusin et al., 2012).

Su impacto no se reduce al ámbito individual, sino que puede consolidarse en prácticas organizativas si no se adoptan medidas explícitas para detectarlos y mitigarlos (Bielby, 2000; Williams, 2014). De ahí la importancia de promover la formación en sesgos, revisar los procesos de toma de decisiones y fomentar entornos donde la conciencia crítica y la equidad sean pilares estructurales (Banaji & Greenwald, 2013; Steele, 2010).

En el siguiente apartado se abordarán otras formas en que los sesgos pueden manifestarse y amplificarse, especialmente cuando se incorporan tecnologías como la IA en los procesos de RR. HH. (Eubanks, 2018; Noble, 2018).

b) Sesgos algorítmicos: cuando los sistemas aprenden nuestras desigualdades

Los sistemas de IA no son ajenos a los sesgos. Muy al contrario, pueden absorberlos, amplificarlos y perpetuarlos si no se diseñan, entrenan y supervisan con criterios de equidad. Lo que comenzó como un sesgo inconsciente en decisiones humanas puede convertirse, al ser trasladado a sistemas algorítmicos, en una forma estructural de discriminación automatizada.

¿Qué es el sesgo algorítmico?

El sesgo algorítmico se refiere a cualquier resultado sistemáticamente injusto o desigual producido por un sistema automatizado. Puede tener múltiples causas y manifestarse en diferentes fases del ciclo de vida del algoritmo: desde la selección de datos hasta el diseño técnico del modelo o su interpretación práctica (Mehrabi et al., 2021).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

A diferencia del sesgo humano, que opera a nivel individual, el sesgo algorítmico puede escalar rápidamente, afectando a miles de personas de forma consistente y silenciosa si no se identifican y corrigen sus fuentes.

Principales fuentes de sesgo en sistemas de IA:

- **Datos históricos con sesgos humanos:** cuando se entrena un modelo con decisiones pasadas discriminatorias, el sistema tiende a replicarlas (O'Neil, 2016).
- **Variables proxy:** atributos aparentemente neutros que están correlacionados con factores sensibles como raza o género, como el código postal o el nivel educativo (Barocas & Selbst, 2016).
- **Desbalance entre clases:** cuando ciertos grupos están poco representados en los datos, el algoritmo aprende peor sobre ellos (Hardt et al., 2016).
- **Diseño de modelos sin criterios de equidad:** optimizar solo por precisión sin medir diferencias en el desempeño por subgrupos (Barocas et al., 2019).
- **Confianza ciega en los resultados:** asumir que el sistema es objetivo solo por ser automatizado [Floridi & Cowls, 2019].

A partir de estas causas, se derivan distintos tipos de sesgos que pueden comprometer la equidad y la fiabilidad de los sistemas automatizados. Identificar con claridad su naturaleza permite intervenir de forma específica en su detección y mitigación.

A continuación, se presentan los sesgos algorítmicos más comunes, acompañados de ejemplos y referencias clave que los han documentado en la literatura.

Eliminar sesgos, ampliar oportunidades

La aplicación de sistemas de IA en procesos de selección, promoción o evaluación del desempeño puede tener un impacto profundo en la empleabilidad de personas en situación de vulnerabilidad. Los sesgos algorítmicos afectan especialmente a colectivos como:

- **Mujeres víctimas de violencia de género,** que pueden ser penalizadas si los datos históricos reflejan trayectorias laborales interrumpidas o precarizadas.
- **Personas migrantes o refugiadas,** cuyos perfiles pueden ser excluidos si el sistema aprende a asociar ciertos nombres, códigos postales o instituciones educativas extranjeras con menor "ajuste cultural" o "desempeño previsto".
- **Jóvenes ex tutelados por la Administración Pública,** que suelen tener historiales fragmentados o sin referencias laborales, lo que puede ser interpretado como una falta de mérito o de compromiso.

Estos sesgos no solo perpetúan desigualdades estructurales, sino que impiden a las organizaciones descubrir talento valioso. La detección y corrección de estos sesgos es una herramienta poderosa para facilitar itinerarios personalizados de inserción sociolaboral más justos y efectivos.

Además del análisis técnico del sesgo algorítmico, resulta crucial reconocer su dimensión social. Cuando un sistema o modelo de inteligencia artificial desestima automáticamente a una candidata por tener lagunas laborales asociadas a situaciones de violencia, o no considera a un joven migrante por no contar con un historial en bases de datos locales, no estamos solo ante un error técnico: estamos

también reforzando barreras de acceso al empleo. Incorporar criterios de equidad e inclusión desde el diseño del sistema es una buena práctica ética y una estrategia de empleabilidad y justicia social.

Tabla 6. *Tipología de sesgos algorítmicos*

Tipo de sesgo	Descripción	Ejemplos	Autores | Fuentes
Representación	Datos que no reflejan la diversidad real	Sistema de reconocimiento facial falla con personas de piel oscura	Buolamwini & Gebru (2018) Barocas et al. (2019)
Exclusión	Eliminar variables relevantes para ciertos grupos	Omitir datos de personas en situación discapacidad en procesos de selección	Mehrabi et al. (2021) O'Neil (2016)
Confirmación	Reproducción de patrones sesgados del pasado	Preferir perfiles masculinos por reflejar contrataciones anteriores	Angwin et al. (2016) Bolukbasi et al. (2016)
Automatización	Asunción de que el algoritmo no comete errores	Aceptar sin revisión un sistema de scoring crediticio	IBM (2018) Floridi & Cowls (2019)
Medición	Errores sistemáticos al recolectar datos	Evaluar desempeño con criterios distintos según el grupo.	Suresh & Guttag (2021)
Selección	Entrenar con muestras no representativas	Usar solo datos de empleados actuales, ignorando rechazados	Mehrabi et al. (2021) Barocas et al. (2019)
Observación	Registro subjetivo o inconsistente de datos	Supervisores califican mejor a ciertos grupos sin justificación	Holstein et al. (2019)
Agregación	Aplicar un solo modelo a subgrupos distintos	Usar un modelo único para evaluar técnicos y perfiles comerciales	Dastin (2018) Chouldechova (2017)
Variables proxy	Uso de variables indirectamente sensibles	Código postal asociado a etnia o nivel socioeconómico	Barocas & Selbst (2016)
Desbalance de clases	Datos escasos de ciertos colectivos	El modelo acierta más con hombres que con mujeres por falta de ejemplos	Hardt et al. (2016) Mehrabi et al. (2021)
Diseño	Decisiones técnicas sin perspectiva de equidad	Optimizar solo por precisión sin tener en cuenta el impacto entre grupos	Barocas et al. (2019)

Fuente: elaboración propia.

Autores clave en el estudio del sesgo algorítmico:

- **Cathy O'Neil**, *Weapons of Math Destruction* (2016): pionera en visibilizar cómo los algoritmos amplifican desigualdades.
- **Joy Buolamwini & Timnit Gebru,** *Gender Shades* (2018): demostraron sesgos raciales y de género en visión artificial.
- **Sorelle A. Friedler, Carlos Scheidegger & Suresh Venkatasubramanian:** desarrollaron marcos computacionales para auditar equidad.
- **Moritz Hardt, Solon Barocas & Arvind Narayanan**, *Fairness and Machine Learning* (2019): visión técnica y crítica del campo.
- **Imane F. Mehrabi et al.**, *A Survey on Bias and Fairness in Machine Learning* (2021): revisión de sesgos y estrategias de mitigación.
- **Sandra Wachter, Brent Mittelstadt & Chris Russell:** abordan la explicabilidad y el marco jurídico de la IA en Europa.

Compromiso ético y acción organizacional

Abordar los sesgos —tanto humanos como algorítmicos— no es solo un desafío técnico, sino una

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

responsabilidad ética y estratégica para cualquier organización que utilice IA en la gestión del talento. Es crucial implementar mecanismos de supervisión, auditoría y revisión continua que permitan identificar y mitigar estos sesgos (Buolamwini & Gebru, 2018; Floridi & Cowls, 2019; Barocas et al., 2019), garantizando procesos más justos, transparentes e inclusivos.

Además, la sensibilización y formación del personal encargado tanto del diseño de sistemas como de la toma de decisiones en RR.HH. es clave para detectar riesgos, corregir desviaciones y fomentar una cultura organizacional consciente de su impacto social.

La transformación digital debe estar al servicio de una gestión del talento más equitativa. Solo mediante una combinación de diseño responsable, vigilancia activa y compromiso crítico será posible garantizar que la IA no perpetúe desigualdades preexistentes, sino que contribuya a construir entornos laborales más diversos, inclusivos y sostenibles.

El análisis técnico y normativo más detallado sobre sesgos será desarrollado en el último apartado de este capítulo, Consideraciones éticas y legales, donde se abordarán marcos regulatorios, directrices de mitigación y estrategias de auditoría algorítmica.

La eliminación o mitigación de estos sesgos algorítmicos no es solo una cuestión técnica: es un compromiso ético con la igualdad de oportunidades en el acceso al empleo. Su identificación y corrección activa contribuye a garantizar procesos de selección más justos y a favorecer la inserción sociolaboral de colectivos especialmente expuestos a la discriminación estructural, como las mujeres víctimas de violencia, las personas migrantes o quienes han vivido bajo tutela institucional.

1.2.6 Impacto de la normativa europea

El despliegue de la IA en los procesos de RR. HH. ha generado una oleada de innovación sin precedentes, pero también ha revelado lagunas normativas importantes. En respuesta, la Unión Europea ha dado un paso decisivo al impulsar un marco regulador que aspira no solo a proteger los derechos fundamentales, sino también a garantizar que la transformación digital sea inclusiva, ética y socialmente sostenible (European Commission, 2020).

Entre las normativas más relevantes se encuentra el RIA, que clasifica los sistemas de IA según el nivel de riesgo que entrañan para los derechos de las personas. En este marco, muchas aplicaciones de IA en RR. HH. —como la selección automatizada, la evaluación algorítmica del desempeño o la asignación de trayectorias profesionales— son consideradas de “riesgo alto” (European Parliament and Council, 2024). Esto implica la necesidad de cumplir con exigencias adicionales de transparencia, trazabilidad, supervisión humana y calidad de los datos (High-Level Expert Group on AI, 2019).

Aunque el AI Act no prohíbe el uso de IA en RR. HH., sí redefine las condiciones bajo las cuales puede aplicarse legítimamente. Las organizaciones deberán demostrar que los algoritmos no reproducen sesgos, que las decisiones pueden explicarse y que existen mecanismos internos para corregir errores (Access Now, 2022; AlgorithmWatch, 2023). Estas exigencias no deben entenderse solo como obligaciones legales, sino como una invitación a rediseñar los procesos desde una perspectiva ética y centrada en la persona (OECD, 2021).

Este enfoque regulador se articula además con otros instrumentos ya en vigor, como el Reglamento General de Protección de Datos, que limita el tratamiento automatizado de datos personales sin

consentimiento explícito o justificación legítima (European Union, 2016). La combinación de ambos marcos impone un cambio de paradigma: los sistemas de IA no pueden ser tratados como herramientas neutras, sino como dispositivos regulados, auditables y sujetos a control (Wachter, Mittelstadt & Floridi, 2017).

Tabla 7: *Requisitos del AI Act aplicables al uso de IA en RR. HH.*

Elemento regulado	Aplicación en RR. HH.	Obligaciones HRAIS
Clasificación como riesgo alto	Reclutamiento automatizado, evaluación de desempeño, ascensos	Sujeto a obligaciones específicas del RIA
Gobernanza del dato	Bases de datos de candidatos, empleados, rendimiento	Control de sesgos, calidad y representatividad de los datos
Transparencia	Algoritmos que afectan a decisiones laborales	Información clara sobre el uso y funcionamiento del sistema
Supervisión humana	Decisiones que afectan a contratación o desarrollo profesional	Supervisión significativa y capacidad de intervención manual
Trazabilidad	Historial de decisiones automatizadas	Registro de operaciones y posibilidad de auditoría
Evaluación de riesgos	Nuevas herramientas de IA para RR. HH.	Evaluación de impacto antes de su puesta en marcha
Documentación técnica	Proveedores o integradores de IA para RR. HH.	Manuales, fichas técnicas, protocolos de uso accesibles y actualizados
Vigilancia posterior al mercado	IA implantada en procesos de personas	Monitoreo de funcionamiento y notificación de incidentes

Fuente: elaboración propia.

Estos marcos regulatorios no solo responden a exigencias técnicas o de transparencia. Tienen como objetivo proteger los derechos fundamentales de todas las personas, en especial de aquellos colectivos más vulnerables, promoviendo un uso de la IA que garantice la no discriminación, el consentimiento informado y la supervisión humana activa en entornos laborales automatizados.

Para los departamentos de RR. HH., esto implica asumir un rol proactivo en el cumplimiento normativo, trabajando en colaboración con áreas técnicas, jurídicas y de auditoría. Se trata no solo de evitar sanciones, sino de posicionar a la función de RR. HH. como garante del uso ético y legal de la tecnología (International Labour Organization, 2021). Esta evolución refuerza el papel estratégico del área de personas, no como usuaria pasiva de soluciones externas, sino como corresponsable de su diseño, implementación y supervisión.

En los próximos capítulos, se abordarán con mayor profundidad los marcos legales europeos y españoles aplicables al uso de la IA en el entorno laboral. El presente subapartado solo ha pretendido esbozar los principales efectos que esta normativa ya está teniendo —y previsiblemente tendrá— en la transformación del área de RR. HH. (Veale & Edwards, 2018; FRA, 2020).

La incorporación de la IA en la gestión de personas está transformando no solo los procesos, sino los fundamentos mismos de la organización del trabajo. Si bien los beneficios son evidentes en términos de eficiencia, personalización y capacidad predictiva, también lo son los riesgos asociados a la

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

opacidad, la desigualdad y la erosión de la confianza. Como se ha mostrado, afrontar estos retos no depende únicamente de la tecnología en sí, sino de las decisiones organizativas, éticas y normativas que la acompañan.

La función de RR. HH. está llamada a desempeñar un papel estratégico en esta transición: como impulsora de talento, garante de la equidad y mediadora entre innovación y responsabilidad. En este contexto, el próximo apartado abordará un pilar clave para una IA justa e inclusiva: la gobernanza del dato. Porque no puede haber IA ética sin datos bien gobernados.

1.3 Gobernanza del dato como pilar de una IA responsable e inclusiva

1.3.1 Introducción: ¿Por qué importa la Gobernanza del dato en el contexto actual?

La Gobernanza del dato en el ámbito laboral no debe concebirse como una función puramente técnica ni relegarse al plano operativo de las tecnologías de la información. En realidad, constituye un pilar estructural de la justicia organizativa y de la rendición de cuentas en la era de la automatización algorítmica. Esto es especialmente relevante cuando la inteligencia artificial interviene en procesos clave de la gestión de personas: desde la selección y evaluación del talento, hasta la formación, promoción o incluso la desvinculación de trabajadores (Jarrahi et al., 2024).

En este escenario, los datos no son solo entradas para sistemas de decisión automatizados, sino representaciones codificadas de trayectorias, competencias, rendimientos y circunstancias humanas. Por tanto, cualquier sesgo, error u opacidad en los datos puede traducirse en injusticias materiales. Un sistema de IA mal alimentado —con datos incompletos, desactualizados, excluyentes o heredados de lógicas discriminatorias— puede reproducir estereotipos y consolidar desigualdades estructurales en lugar de mitigarlas (AI and Ethics, 2024).

El despliegue de una Gobernanza del dato orientada a la equidad no solo fortalece la transparencia organizativa, sino que responde a compromisos más amplios de justicia social y desarrollo sostenible. En línea con el Eje 1, referenciado en el prólogo de este Libro Blanco, la garantía de datos fiables, trazables y no discriminatorios resulta clave para promover la autonomía personal y la inclusión social de colectivos históricamente excluidos del mercado laboral, como personas con discapacidad, mujeres víctimas de violencia o jóvenes ex tutelados.

Asimismo, desde el Eje 3, el uso responsable de datos se vincula a los derechos digitales y la reducción de la brecha digital, haciendo accesible la IA a toda la población laboral, incluso aquella más vulnerable o con menor nivel de digitalización. Finalmente, en sintonía con el Eje 5, una gobernanza bien diseñada permite detectar y corregir sesgos algorítmicos que afectan a personas migrantes, refugiadas o en riesgo de exclusión, contribuyendo a su empleabilidad y plena inserción sociolaboral.

La creciente delegación de decisiones a sistemas algorítmicos exige que nos interroguemos no solo sobre la capacidad técnica de la IA, sino también sobre la legitimidad de los datos que sustentan sus recomendaciones. Este enfoque ha sido recogido explícitamente en el Reglamento europeo de Inteligencia Artificial, que en su artículo 10 establece obligaciones para asegurar la calidad, trazabilidad y representatividad de los datos en sistemas de alto riesgo, incluidos los aplicados en recursos humanos (AI Act, art. 10).

La Gobernanza del dato permite establecer mecanismos de control para que estos datos sean:

- **Confiables**: validados, completos y actualizados.
- **Trazables**: con origen, transformación y uso documentados.
- **Auditables**: disponibles para revisiones internas o externas.
- **Representativos**: reflejando con fidelidad la diversidad del entorno laboral (Taylor Wessing, 2024).

Además, un marco de gobernanza robusto facilita el ejercicio del control humano significativo, principio clave para la supervisión ética de la IA en el entorno de trabajo (Springer, 2024). Solo si los datos están estructurados y accesibles, es posible intervenir, cuestionar o modificar una decisión basada en IA, protegiendo así los derechos fundamentales de los trabajadores.

Por tanto, no hay IA responsable sin Gobernanza del dato. La gobernanza es la infraestructura técnica, organizativa y ética que permite construir sistemas de IA que no solo funcionen bien, sino que funcionen justamente. Constituye la base sobre la que se edifica la confianza de empleados, empleadores, reguladores y sociedad civil en los sistemas de gestión algorítmica del talento (Sngular, 2023). Y es, en última instancia, una condición de posibilidad para una IA que respete la dignidad humana, fomente la inclusión y refuerce la empleabilidad en lugar de erosionarla (2Brains, 2024).

1.3.2 Marco conceptual: Gobernanza del dato y sus dimensiones

La Gobernanza del dato en el contexto de la inteligencia artificial aplicada a los recursos humanos no puede limitarse a una serie de normas administrativas o procedimientos técnicos. Se trata de un marco estratégico e integrador que abarca dimensiones éticas, legales, organizativas y técnicas, y que debe situarse en el núcleo de cualquier iniciativa de automatización de procesos laborales (Raisch & Krakowski, 2022).

A continuación, se desarrollan las cuatro dimensiones clave de este marco conceptual:

a) Dimensión ética: equidad, inclusión y autonomía

La dimensión ética de la Gobernanza del dato exige que los sistemas de IA respeten valores fundamentales como la dignidad humana, la no discriminación, la equidad y la inclusión. En el ámbito laboral, esto se traduce en:

- Evitar el uso de datos que refuercen **sesgos estructurales,** como los derivados de algún antecedente histórico discriminatorio (Jarrahi et al., 2024).
- La necesidad de diseñar políticas de datos que aseguren **representatividad y diversidad**, aspectos clave para no excluir colectivos históricamente marginados (Jobin, Ienca & Vayena, 2019).
- Fomentar la **explicabilidad de las decisiones algorítmicas**, como principio rector de una IA transparente y humanamente comprensible (HR.com, 2024).

b) Dimensión legal: Cumplimiento normativo y derechos fundamentales

Desde una perspectiva legal, la Gobernanza del dato debe garantizar el cumplimiento de marcos normativos como el GDPR o el RIA, que clasifican como “alto riesgo” los sistemas aplicados en la

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030 Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

gestión de personal. Esta clasificación exige mayor rigor en la calidad de los datos, el control humano y la documentación de decisiones (AI in HR: The Current Legal Landscape, 2024).

Además, el respeto a los derechos fundamentales de los trabajadores —como el acceso, la corrección o la oposición al tratamiento de sus datos— debe ser garantizado a través de prácticas claras y verificables. Esto incluye también evaluaciones de impacto legal y ético antes del despliegue de cualquier sistema de IA en RR. HH. (Legal Resolve, 2025).

c) Dimensión organizativa: gobernanza institucional y cultura del dato

La dimensión organizativa requiere una infraestructura interna que combine especialización técnica, ética y legal. Muchas organizaciones están avanzando hacia la creación de comités de revisión algorítmica o charters éticos que guían el desarrollo y uso responsable de la IA (Morley et al., 2024; AMS Charter, 2024).

Asimismo, fomentar una cultura organizativa basada en la responsabilidad de los datos no solo mejora la calidad de los sistemas, sino que refuerza la confianza de empleados y stakeholders. Este proceso requiere liderazgo comprometido y formación transversal en los valores y riesgos del uso de IA (Osborne Clarke, 2024).

d) Dimensión técnica: calidad, trazabilidad y seguridad de los datos

Los sistemas de IA solo pueden operar de forma legítima si están alimentados por datos gobernados con rigor técnico. Esto implica asegurar:

- **Calidad de los datos** a través de validación constante y eliminación de errores sistemáticos.
- **Trazabilidad completa,** que permita auditar decisiones automatizadas y atribuir responsabilidades si fuera necesario (European AI Alliance, 2024).
- **Seguridad y anonimización,** como salvaguardas esenciales frente a brechas de privacidad o usos no autorizados.

El uso de entornos de prueba controlados (sandbox ético) se está consolidando como práctica recomendada para evaluar impactos antes del despliegue final (LinkedIn, 2024).

Este marco conceptual no solo responde a exigencias técnicas o normativas, sino que se alinea directamente con los objetivos sociales y de inclusión planteados en los tres ejes estratégicos del proyecto.

Eje 1, una gobernanza que garantice datos equitativos, trazables y auditables contribuye a derribar barreras de acceso al empleo para personas en situación de vulnerabilidad: víctimas de violencia de género, personas con discapacidad o jóvenes sin red familiar de apoyo.
En relación con el Eje 3, asegurar estándares de gobernanza técnica y explicabilidad promueve el ejercicio de derechos digitales y la reducción de brechas de acceso a la tecnología, facilitando el uso seguro y comprensible de la IA en entornos laborales cada vez más automatizados.
Finalmente, conforme al Eje 5, cada dimensión del marco —ética, legal, organizativa y técnica— contribuye a mejorar la empleabilidad de colectivos marginados mediante sistemas que puedan ser auditados, comprendidos y corregidos si reproducen desigualdades.

Las cuatro dimensiones de la gobernanza del dato —ética, legal, organizativa y técnica— deben integrarse como capas interdependientes de un mismo sistema. Su aplicación coordinada es la única vía para garantizar que la IA aplicada a RR. HH. no solo sea eficiente, sino también legítima, justa y socialmente aceptable (IBM, 2024).

1.3.3 Enfoque estratégico: Políticas de datos para una IA justa

En el marco de una inteligencia artificial aplicada al ámbito laboral, la Gobernanza del dato debe dejar de ser una tarea periférica o reactiva para convertirse en una estrategia central de transformación organizativa. Las empresas no solo deben adoptar tecnologías avanzadas, sino también garantizar que los sistemas que diseñan, implementan o contratan lo hagan sobre la base de principios sólidos de justicia algorítmica, transparencia e inclusión (Frisby, 2024).

Esto exige una integración deliberada de políticas de datos dentro de las agendas de transformación digital, cultural y de gestión del talento. A continuación, se describen cuatro líneas estratégicas prioritarias:

a) Diseño de políticas de recopilación y uso de datos laborales

Una IA justa comienza con la recolección responsable de datos. Las organizaciones deben establecer políticas que:

- **Delimiten claramente qué datos pueden recopilarse** en función de su relevancia laboral, descartando variables no justificadas (como creencias religiosas o afiliaciones políticas), de acuerdo con buenas prácticas establecidas por marcos como el GDPR y recomendaciones sectoriales (HRMagazine, 2024).
- **Eviten la inclusión de atributos sensibles** que introduzcan sesgos directos o indirectos.
- **Establezcan principios de finalidad y proporcionalidad**, permitiendo a los trabajadores oponerse al uso de determinados datos, según el contexto.

b) Auditorías de sesgos algorítmicos y evaluaciones de impacto ético

Ninguna política de datos puede ser considerada responsable si no incluye mecanismos activos de vigilancia y mejora continua. Para ello es fundamental:

- Realizar **auditorías periódicas de sesgos algorítmicos,** apoyándose en metodologías como las de Warden AI, herramienta promovida en Reino Unido para auditorías continuas en tecnologías de RR. HH. (UK Gov, 2024).
- Aplicar evaluaciones de impacto ético que anticipen efectos adversos antes del despliegue de sistemas automatizados (Optiblack, 2024).
- Integrar estos procesos con participación plural: comités interdisciplinarios y representantes sindicales.

Estas prácticas permiten construir un entorno de confianza organizativa, mitigando riesgos técnicos y sociales (HR Dive, 2024).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

c) Procesos de anonimización y minimización de datos

El tratamiento de datos personales debe seguir principios de minimización y anonimización efectiva, especialmente en procesos exploratorios o de validación de modelos. Para ello:

- Se recomienda implementar **tecnologías de anonimización** impulsadas por IA, como las empleadas en encuestas internas de clima laboral y cultura (DataGuard, 2024).
- Además, se deben establecer **ventanas temporales claras para conservación**, asegurando la eliminación o anonimización irreversible de los datos tras su uso.

Estas prácticas refuerzan tanto la seguridad de los datos como el cumplimiento normativo proactivo, al alinearse con métodos reconocidos de gobernanza técnica (Restack, 2024).

d) Estándares de explicabilidad algorítmica

La explicabilidad no es un lujo técnico, sino una obligación legal y social, reconocida por el AI Act de la UE (Springer, 2024). Para cumplir con este principio, las organizaciones deben:

- Establecer **estándares de explicabilidad diferenciados por nivel de audiencia,** incluyendo a trabajadores y comités sindicales (Mayer Brown, 2024).
- Adoptar **buenas prácticas sectoriales** como las desarrolladas por empresas líderes en RR. HH., por ejemplo, el AI Explainability Statement de HireVue (HireVue, 2024).

Una explicabilidad bien implementada permite el control humano significativo y refuerza la trazabilidad y legitimidad de las decisiones algorítmicas en entornos laborales automatizados.

Las políticas de datos descritas en este enfoque estratégico se alinean de forma directa con los principios rectores de los ejes estratégicos del proyecto IA+Igual. En consonancia con el Eje 1, la correcta delimitación de los datos recogidos —excluyendo variables sensibles no justificadas— permite eliminar obstáculos de acceso al empleo para personas en riesgo de exclusión.

A su vez, la promoción de estándares de explicabilidad y transparencia activa responde al Eje 3, favoreciendo el ejercicio efectivo de derechos digitales por parte de toda la población laboral, incluso la más alejada de los entornos tecnológicos. Por último, prácticas como la auditoría de sesgos y la minimización de datos redundan en el Eje 5, contribuyendo a mejorar la empleabilidad y los itinerarios de inserción sociolaboral de colectivos vulnerables, mediante el desarrollo de sistemas más justos y reparadores.

Una IA justa en el ámbito laboral solo es posible si se construye sobre bases políticas claras y operativas, que combinen visión ética con ejecución técnica y compromiso institucional. Las políticas de datos no deben improvisarse, sino diseñarse para que la transformación digital se convierta en una palanca de equidad y no en un nuevo vector de exclusión (Gohire, 2024).

1.3.4 Gobernanza transversal: Inclusión y diversidad como principios guía

Una Gobernanza de datos verdaderamente responsable no puede limitarse al cumplimiento normativo o a la eficiencia técnica. Debe incluir, de forma explícita, el compromiso con la inclusión y la diversidad como ejes centrales del diseño, implementación y supervisión de sistemas de inteligencia

artificial en el entorno laboral. Esta visión está en línea con el enfoque emergente de "IA para todos", que promueve modelos más inclusivos desde su concepción (AI for All, 2023).

Cuando los algoritmos se entrenan, despliegan y evalúan sin una perspectiva inclusiva, corren el riesgo de amplificar brechas estructurales preexistentes. La Gobernanza del dato, por tanto, debe incorporar mecanismos para prevenir estas distorsiones y asegurar que la automatización del trabajo sirva a todos los colectivos, y no solo a una mayoría estadística (Springer, 2023).

a) Representación equitativa en los datos de entrenamiento

Los algoritmos de IA en RR. HH. solo serán justos si se entrenan con datos que reflejen la diversidad real de la fuerza laboral. Como señalan investigaciones recientes, la falta de diversidad en los datos de entrenamiento compromete gravemente la equidad de los resultados algorítmicos (SHRM, 2023).

- Se debe **garantizar la inclusión de perfiles diversos,** incluso de grupos históricamente subrepresentados.
- Es clave revisar los datos históricos con una **mirada crítica**, ya que pueden incorporar sesgos institucionales pasados.
- Además, los métodos de validación deben contemplar la **interseccionalidad**, para identificar sesgos combinados de género, edad, origen o discapacidad (AIHR Institute, 2023).

b) Participación activa y plural en la definición de reglas

Una gobernanza inclusiva no se logra únicamente mediante ingeniería de datos: requiere abrir procesos de decisión a todos los actores implicados. Diversos estudios apuntan a que la legitimidad social de la IA en entornos laborales aumenta significativamente cuando los propios trabajadores participan en la construcción de las reglas algorítmicas (JAIM, 2023).

- Es crucial involucrar a los **sindicatos, comités éticos y representantes legales** desde la fase de diseño.
- La participación debe ser **estructurada y deliberativa,** no meramente consultiva.
- Así se fomenta una **cultura de corresponsabilidad tecnológica**, donde el despliegue de IA se acompaña de legitimidad organizativa.

c) Monitoreo continuo de impactos diferenciados

La equidad no puede garantizarse sin una vigilancia continua. Como señalan prácticas destacadas en auditoría algorítmica, muchos efectos adversos no se detectan en las pruebas iniciales, sino durante el uso sostenido de los sistemas (HR Dive, 2024).

- Las organizaciones deben implementar indicadores de impacto por grupo, revisados periódicamente.
- No solo se trata de métricas cuantitativas, sino también de mecanismos cualitativos que capten percepciones de justicia y discriminación.
- Ante resultados adversos, deben existir rutas claras de corrección y mecanismos de rendición de cuentas.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

La adopción de una Gobernanza del dato centrada en la inclusión y la diversidad se convierte en el principal vehículo para dar respuesta concreta a los objetivos planteados por los ejes 1, 3 y 5. Desde el Eje 1, garantizar que los datos representen fielmente a todos los colectivos —y no solo a las mayorías históricas— implica abrir oportunidades reales de inserción laboral para personas con discapacidad, víctimas de violencia o jóvenes sin redes de apoyo.

A su vez, el Eje 3 exige que los sistemas sean no solo técnicamente sólidos, sino también comprensibles y accesibles, especialmente para quienes se encuentran más alejados de la digitalización. Finalmente, el Eje 5 encuentra en el monitoreo y corrección de impactos discriminatorios una herramienta clave para romper los ciclos de desigualdad y construir trayectorias de empleabilidad estables para grupos tradicionalmente marginados.

Una IA verdaderamente inclusiva en el trabajo requiere algo más que datos limpios o decisiones transparentes: exige un compromiso organizativo firme con la equidad y la diversidad. Este compromiso debe estar sustentado por una Gobernanza del dato que incluya representación, participación y vigilancia activa, como ya están recomendando las mejores prácticas internacionales (LinkedIn, 2023).

1.3.5 Recomendaciones prácticas

La construcción de una inteligencia artificial responsable en el ámbito laboral no puede limitarse a principios abstractos. Requiere acciones concretas, replicables y evaluables que operativicen la Gobernanza del dato. Estas recomendaciones se apoyan en una sólida literatura sobre ética aplicada, gobernanza algorítmica y participación institucional en IA (Jobin et al., 2024).

a) Establecer un código ético del dato

El primer paso es establecer un código ético del dato que guíe el tratamiento responsable de la información a lo largo de su ciclo de vida. Tal como recomiendan organismos internacionales y consultoras especializadas, estos códigos deben ser operativos, revisables y específicos al sector de aplicación (EY, 2024).

- Deben abordar principios como minimización, finalidad y no discriminación.
- Tienen que incluir criterios para evitar el uso de datos sensibles no justificados.
- Es necesario prever mecanismos de revisión periódica con base en incidentes y auditorías internas.

b) Implementar mecanismos de supervisión algorítmica con participación plural

Toda gobernanza efectiva requiere supervisión externa al sistema. En IA, esto se traduce en la creación de comités de revisión algorítmica que cuenten con perfiles diversos: técnicos, legales, sindicales y de RR. HH. Este modelo ya se está explorando en múltiples organizaciones como mecanismo de control y evaluación (Morley et al., 2024).

- Estas juntas deben tener acceso a la lógica de los modelos.
- Es necesario que sean consultadas antes del despliegue de sistemas de alto impacto.
- Hace falta autoridad para exigir auditorías o paralizar algoritmos discriminatorios.

c) Crear espacios deliberativos con actores diversos

Más allá de los comités formales, se deben crear espacios deliberativos que promuevan la corresponsabilidad en el uso de IA. Se ha demostrado que la deliberación estructurada, incluso entre humanos y sistemas de IA, mejora la calidad y legitimidad de las decisiones (Starke et al., 2024).

- Pueden adoptar la forma de foros internos, grupos de enfoque o laboratorios de innovación.
- Deben incorporar a trabajadores, expertos, dirección y actores externos relevantes.
- Sus propuestas deben integrarse en la política institucional y no quedar en el plano simbólico (West et al., 2023).

d) Priorizar la transparencia proactiva

La transparencia debe ser proactiva, accesible y adaptada al nivel de comprensión de los públicos involucrados. Esto implica no solo cumplir formalmente con la explicabilidad exigida por el AI Act, sino comunicar de manera clara y útil (Springer, 2024).

- Crear portales donde los trabajadores puedan consultar cómo se toman decisiones algorítmicas que les afectan.
- Publicar informes de impacto, auditorías y revisiones de sesgo en lenguaje comprensible (LinkedIn, 2024).
- Establecer vías claras para solicitar explicaciones, corregir decisiones y ejercer derechos.

Estas recomendaciones convierten principios abstractos en acciones específicas, evaluables y adaptables a distintos contextos laborales. Reflejan el consenso emergente de que una IA ética y eficaz solo es posible si se integra la Gobernanza del dato con mecanismos institucionales de revisión, participación y transparencia (Workable, 2023).

Cada una de las recomendaciones aquí propuestas constituye un mecanismo directo para operativizar los compromisos asumidos en los ejes 1, 3 y 5 del proyecto. El desarrollo de códigos éticos y auditorías con perspectiva de derechos favorece el empoderamiento de colectivos vulnerables y la prevención de exclusión laboral injustificada -Eje 1-.

La creación de espacios deliberativos y portales de explicabilidad mejora el acceso a información comprensible y fortalece los derechos digitales y la alfabetización tecnológica de trabajadores con menos recursos o experiencia digital -Eje 3-. Finalmente, la supervisión plural y la corrección de sesgos permiten diseñar sistemas que apoyen de forma activa la empleabilidad y los procesos de inserción sociolaboral de personas migrantes, refugiadas o en situación de vulnerabilidad estructural -Eje 5-.

e) Proyección futura

La gobernanza del dato en el ámbito laboral debe ser concebida no como un marco estático, sino como una arquitectura dinámica, en evolución constante, capaz de adaptarse a nuevos retos tecnológicos, normativos y sociales.

La irrupción de modelos de inteligencia artificial generativa, con capacidad de generar contenidos, evaluaciones o descripciones de candidatos en tiempo real, plantea desafíos inéditos en materia de

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

trazabilidad, verificabilidad y sesgo emergente. Estos sistemas, por su complejidad y autonomía, exigen reforzar los principios de explicabilidad y control humano significativo que sustentan todo marco de gobernanza responsable.

Junto a ello, se hace indispensable una apuesta firme por la alfabetización en inteligencia artificial en los entornos laborales. Comprender cómo operan los algoritmos, qué decisiones pueden automatizar y cómo afectan los derechos individuales, debe convertirse en una competencia básica para empleados, responsables de RR. HH. y agentes sociales. La gobernanza no solo necesita políticas y estructuras, sino también personas capacitadas para entender, aplicar y cuestionar la IA en el trabajo.

Finalmente, el despliegue efectivo de este modelo de gobernanza exige un rol activo de los poderes públicos: promoviendo marcos regulatorios claros, financiando la investigación aplicada, impulsando mecanismos de evaluación algorítmica y garantizando que las tecnologías no profundicen desigualdades preexistentes.

Solo una Gobernanza del dato que actúe como garante de la equidad, la explicabilidad y la participación permitirá que la inteligencia artificial en el ámbito laboral sea, verdaderamente, una herramienta de emancipación y no de exclusión.

1.4 IA aplicada a diferentes procesos en RR. HH.

La IA está transformando profundamente la función de RR. HH. en las organizaciones contemporáneas. Su implementación no se limita a automatizar tareas operativas: permite reconfigurar procesos estratégicos vinculados a la atracción, desarrollo, bienestar y desvinculación del talento, por ejemplo. Esta capacidad para generar eficiencia, personalización y análisis predictivo ha elevado el papel de la IA como palanca estructural en la gestión del capital humano (AIHR, 2023; IBM, 2023).

Sin embargo, esta transformación no es neutral ni exenta de dilemas. La forma en que se diseña aplica y supervisa la IA en cada proceso de RR. HH. puede ampliar la justicia organizativa y la equidad en el acceso a oportunidades o, por el contrario, reproducir sesgos históricos y profundizar dinámicas de exclusión (Ajunwa, 2020; Eubanks, 2018).

Automatizar tareas dentro de un proceso de toma de decisiones implica también trasladar al sistema una jerarquía de prioridades y, con ello, una determinada visión del talento, la productividad o la normalidad (Barocas, Hardt & Narayanan, 2019).

Este apartado examina cómo se materializa la IA en distintos procesos del ciclo de vida laboral. Lo hace desde una perspectiva crítica, que pone en diálogo el potencial transformador de la tecnología con los riesgos éticos, sociales y organizativos que emergen cuando su aplicación carece de supervisión significativa (European Commission, 2021; Holstein et al., 2019). En cada subapartado se analizarán los beneficios esperados, las alertas principales y las condiciones para una implementación justa, con referencias al marco de gobernanza descrito previamente y una atención especial a su impacto sobre la diversidad, la autonomía y la inclusión laboral.

1.4.1 Gestión del talento: IA en todas las fases del ciclo de vida profesional

La gestión del talento ha sido históricamente una función estratégica dentro de los departamentos de RR. HH., pero la introducción de la IA redefine por completo su alcance, sus tiempos y sus criterios

de decisión. La IA permite analizar trayectorias profesionales, anticipar necesidades de desarrollo, personalizar itinerarios formativos y reducir tiempos de respuesta en procesos clave (AIHR, 2023; McKinsey & Company, 2023). Sin embargo, esta capacidad de intervención a lo largo de todo el ciclo de vida profesional —desde la atracción del talento hasta la desvinculación— exige una reflexión crítica sobre el modo en que se automatizan procesos de valoración y recomendación, así como sobre el papel que desempeña la supervisión humana en las decisiones que afectan directamente al recorrido laboral de las personas (OECD, 2023), tal como establece el Reglamento Europeo de IA.

En lugar de concebir la IA como una capa técnica añadida a los procesos existentes, debe entenderse como un vector estructural que los reconfigura. Su introducción transforma no solo la forma de seleccionar o evaluar, sino también los criterios implícitos sobre lo que se puede considerar un perfil válido, una carrera deseable o una salida justificada (Bersin, 2022). De ahí la necesidad de implementar mecanismos de control humano, validación interseccional y transparencia explicativa en cada fase (Holstein et al., 2019; IBM, 2023).

Asimismo, es esencial tener en cuenta los riesgos de reproducción de desigualdades interseccionales en la aplicación de estos sistemas. Variables como la edad, el género, el origen étnico, la discapacidad o la trayectoria profesional con discontinuidades o no convencional pueden verse penalizadas por modelos entrenados con datos históricos no neutrales. Sin medidas correctoras explícitas, la IA podría reforzar las barreras preexistentes en el acceso, la promoción o la retribución, incluso bajo una apariencia de objetividad.

A continuación, se detallan las principales aplicaciones de la IA en la gestión del talento, así como los beneficios y riesgos asociados en cada una de ellas:

Tabla 8: *Resumen de aplicaciones de IA en la gestión del talento*

Fase del ciclo de vida	Aplicación de IA	Beneficios potenciales	Riesgos y alertas
Atracción del talento	Matching algorítmico de CV Análisis predictivo de desempeño	Eficiencia en el filtrado Ampliación de perfiles afines	Exclusión de trayectorias atípicas Sesgos en datos históricos
Onboarding	Personalización itinerarios Bots de bienvenida IA conversacional	Integración rápida y adaptada a cada rol	Experiencias impersonales Desajuste con expectativas culturales
Desarrollo y carrera	Recomendación de itinerarios profesionales Análisis de potencial Evaluación automatizada del desempeño	Visibilización de talento interno Proyección de carrera Alineación desempeño y oportunidades	Reproducción de techos de cristal Penalización de carreras no lineales Sesgos en la evaluación
Itinerarios formativos	Recomendación personalizada contenidos Sistemas de tutoría virtual adaptativa Análisis automatizado de progreso	Formación alineada perfil y ritmo empleado Identificación temprana necesidades de aprendizaje	No visibilización de necesidades imprevistas Sesgos en el seguimiento del progreso Dificultad para evaluar transferencia real al puesto

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Productividad	Monitorización de tareas Análisis de eficiencia por procesos Predicción de cuellos de botella	Optimización de recursos Identificación de mejoras operativas Toma de decisiones basada en datos	Reducción del trabajo a métricas cuantitativas Vigilancia excesiva Presión constante sin considerar el contexto
Absentismo	Detección de patrones de ausencias Análisis predictivo de factores de riesgo Correlación con condiciones de trabajo	Prevención proactiva Identificación de causas organizativas Diseño de intervenciones tempranas	Estigmatización de perfiles Decisiones preventivas injustificadas Confusión entre causas personales y estructurales
Planes de sucesión	Predicción de relevo generacional Análisis de liderazgo emergente	Anticipación y continuidad organizativa	Favorecimiento de perfiles homogéneos Opacidad en criterios
Rotación y offboarding	Predicción de rotación no deseada Análisis de causas de salida Modelado de abandono Retroalimentación automatizada	Mejora de retención Rediseño organizativo Anticipación de problemáticas estructurales	Uso invasivo de datos sensibles Decisiones preventivas injustificadas Feedback mal interpretado o descontextualizado

Fuente: elaboración propia.

Cada una de estas fases presenta oportunidades reales de mejora, pero también puntos críticos que requieren gobernanza algorítmica, participación sindical y auditoría regular. Todo ello debe enmarcarse en una cultura organizativa del dato y, aún más, en una cultura de gobernanza de la IA, capaz de anticipar impactos, promover usos justos y alinear la tecnología con los valores de la organización. En particular, los sistemas de recomendación aplicados al desarrollo profesional o los modelos de predicción de potencial deben incorporar evaluaciones diferenciadas por grupo y no aplicar métricas homogéneas que invisibilicen la diversidad de trayectorias (Barocas, Hardt & Narayanan, 2019; Ajunwa, 2020). Una IA bien diseñada puede ampliar el acceso al empleo para colectivos en riesgo de exclusión, garantizando una evaluación basada únicamente en la competencia para el puesto (Eje 1 del proyecto). Sin embargo, la equidad algorítmica no basta si no se acompaña de condiciones organizativas capaces de responder a las realidades específicas de estos colectivos. Adaptar puestos a personas con discapacidad o preservar la seguridad y la confidencialidad en casos de violencia de género plantea desafíos que no pueden resolverse únicamente desde el diseño técnico del sistema, sino que exigen sensibilidad institucional, acompañamiento personalizado y marcos de protección reforzada.

La eliminación de sesgos inconscientes en los sistemas de evaluación automatizada mejora directamente la empleabilidad de colectivos tradicionalmente marginados, como migrantes, jóvenes tutelados o mujeres víctimas de violencia, tal como establece el Eje 5 del proyecto. Estas personas suelen enfrentarse a modelos entrenados con datos que penalizan trayectorias discontinuas, experiencias informales o situaciones vitales no normativas. Una IA comprometida con la equidad debe ser capaz de identificar el talento sin dejarse guiar por proxies culturales, lingüísticos o institucionales que operan como filtros indirectos. Para ello, se requieren criterios ajustados, supervisión humana significativa y mecanismos de revisión que permitan detectar y corregir

exclusiones invisibles. Solo así se puede convertir la automatización en una aliada de la inclusión sociolaboral real.

Asimismo, los procesos de compensación y sucesión requieren una atención específica: si bien la IA puede detectar patrones de inequidad que pasan desapercibidos en una revisión manual, también puede consolidar criterios salariales injustos si los datos de referencia no han sido revisados con perspectiva de género, edad o condición socioeconómica (LinkedIn Talent Solutions, 2024; Eubanks, 2018).

Gestionar el talento con apoyo de IA exige algo más que eficiencia: requiere una visión de justicia organizativa que atraviese los algoritmos, las interfaces y las decisiones. Esta visión debe nutrirse tanto de evidencias técnicas como de deliberación colectiva, asegurando que la automatización no reduzca la pluralidad del talento a un conjunto limitado de trayectorias normativas (European Commission, 2021).

1.4.2 Cultura, clima y experiencia del empleado: IA como sensor organizativo

El ambiente laboral ha dejado de concebirse como una dimensión abstracta o anecdótica dentro de las organizaciones. Hoy constituye una variable estratégica con impacto directo sobre la productividad, la retención del talento y el bienestar de las personas trabajadoras (Gartner, 2023; OECD, 2023). En este contexto, la IA se consolida como un instrumento emergente para analizar, mapear y comprender fenómenos organizativos que antes eran difícilmente visibles o medibles (Deloitte, 2023).

Sistemas algorítmicos de análisis de sentimiento, procesadores de lenguaje natural aplicados a encuestas internas, sensores digitales en herramientas colaborativas o modelos predictivos de burnout permiten identificar patrones sutiles de malestar, desconexión emocional o deterioro de la confianza (Microsoft WorkLab, 2023; Calvard, 2021). Estas tecnologías convierten la experiencia del empleado en un conjunto de señales rastreables, abriendo nuevas posibilidades para el diseño de intervenciones proactivas en materia de clima laboral (Springer, 2022).

Es importante distinguir entre los modelos de IA diseñados para anticipar fenómenos estructurales —como el clima organizativo general— y aquellos orientados a detectar señales tempranas de malestar psicosocial en la plantilla. Aunque sus lógicas e impactos pueden diferir, ambos enfoques deben operar desde un análisis agregado, especialmente en el ámbito de la salud mental, donde cualquier intento de personalización directa plantea serios desafíos éticos y de privacidad.

Sin embargo, esta capacidad diagnóstica debe ser acompañada de garantías sólidas que protejan la autonomía individual y los derechos colectivos. La cultura organizativa no puede ser tratada como una fuente de datos neutrales ni como un campo de experimentación sin límites. Es imprescindible asegurar:

- **Transparencia** en el tratamiento, así como en el análisis y uso de los datos generados (IBM, 2023).
- **Consentimiento informado** de los trabajadores, más allá de la mera aceptación contractual (ILO, 2023).
- **Límites claros** al uso de estos sistemas con fines disciplinarios, de control o segmentación impropia (SHRM, 2023; De Stefano et al., 2022).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Una IA aplicada al análisis del clima laboral sin controles éticos adecuados puede derivar en formas encubiertas de vigilancia emocional, socavando la confianza y fomentando dinámicas de autocensura o alienación. Esta preocupación se agrava en los casos en que dicha inferencia emocional se basa en datos biométricos, una práctica expresamente prohibida por el Reglamento de IA (RIA) de la Unión Europea.

Por el contrario, cuando estos sistemas son diseñados e implementados con principios de corresponsabilidad, protección de datos y supervisión plural, pueden convertirse en una infraestructura positiva para el cuidado organizativo. La clave no reside en la tecnología en sí, sino en el marco ético, legal y participativo que la acompaña.

En contextos bien regulados, estas herramientas permiten detectar conflictos latentes antes de que escalen, adaptar las políticas de bienestar a los perfiles reales de la plantilla y promover una cultura del feedback que no dependa exclusivamente de canales jerárquicos (Gartner, 2023; Springer, 2022).

Si se dan las condiciones adecuadas, este uso ético de la IA como sensor organizativo puede alinearse con una visión contemporánea de la cultura empresarial: más horizontal, más participativa y orientada al bienestar estructural.

Sin embargo, asumir este papel implica riesgos importantes si no se establece un marco ético riguroso, ya que una recogida automatizada de señales emocionales o de malestar puede fácilmente derivar en dinámicas de control o autocensura. No se trata de cuantificar el clima, sino de escuchar activamente lo que sucede en el tejido emocional de la organización, siempre desde un enfoque respetuoso, proporcional y abierto al diálogo. Estos sistemas deben servir para traducir hallazgos en decisiones justas, compartidas y sostenibles, no para reforzar asimetrías de poder o limitar la expresión genuina.

Este enfoque requiere también una coordinación estrecha con las estrategias de prevención de riesgos laborales, especialmente en lo relativo al bienestar, como se abordará en el subapartado 1.4.5, donde se profundiza en el rol de la IA en la anticipación de riesgos psicosociales.

1.4.3 Comunicación interna: Personalización sin opacidad

La comunicación interna cumple una función esencial en la cohesión de las organizaciones, especialmente en contextos de cambio, crecimiento o transformación digital. A medida que los entornos laborales se diversifican —en roles, generaciones, niveles de autonomía o ubicación geográfica—, la necesidad de adaptar mensajes, canales y formatos se vuelve más urgente. En este escenario, la IA ofrece herramientas para personalizar la experiencia comunicativa, segmentar contenidos y evaluar el impacto de cada mensaje en tiempo real (McKinsey & Company, 2023; IBM, 2023).

Los sistemas de IA permiten analizar patrones de interacción, preferencias lingüísticas y comportamiento digital para ajustar el contenido, el tono y el canal de comunicación más adecuado para cada perfil. Esto mejora la comprensión y reduce la sobrecarga informativa, aumentando la eficacia global de los flujos internos. Sin embargo, esta capacidad de adaptación conlleva riesgos éticos y organizativos que no deben subestimarse (Harvard Business Review, 2022; Mozilla Foundation, 2022).

La personalización algorítmica en comunicación interna exige una reflexión sobre estos puntos:

• **Evitar la segmentación opaca o invasiva,** en la que los trabajadores desconozcan los criterios por los cuales reciben —o dejan de recibir— ciertos mensajes (Mozilla Foundation, 2022).

• **Preservar la diversidad de voces y estilos comunicativos,** evitando que la eficiencia técnica silencie la participación o la espontaneidad organizativa (Deloitte, 2022).

• **Asegurar la inclusión digital,** especialmente entre colectivos con menor familiaridad tecnológica, discapacidad o acceso desigual a herramientas digitales (SHRM, 2023; Nielsen Norman Group, 2023). Este enfoque se alinea con el Eje 3 del proyecto, que promueve la reducción de la brecha digital y la defensa de los derechos digitales en el contexto laboral, garantizando que la transformación tecnológica sea accesible, comprensible y justa para todas las personas.

Una comunicación interna hipersegmentada, no explicable o basada en métricas de rendimiento puede deteriorar la confianza, amplificar rumores o generar nuevas formas de desinformación organizacional (Edelman, 2023; CIPD, 2023). Además, se corre el riesgo de reforzar dinámicas de exclusión si ciertos grupos quedan sistemáticamente invisibilizados en los canales prioritarios. Un riesgo adicional, cada vez más relevante, es la despersonalización del lenguaje generado por IA: si los mensajes pierden tono humano o credibilidad, pueden ser percibidos como impersonales o artificiales, debilitando el vínculo emocional con la organización.

Por el contrario, cuando se utiliza con responsabilidad, la IA puede contribuir a construir una comunicación interna más equitativa, sensible y participativa. Esto requiere transparencia en los sistemas de recomendación de contenidos, posibilidad de retroalimentación activa y mecanismos de supervisión humana que velen por la coherencia narrativa de la organización (European Commission, 2021; Harvard Business Review, 2022).

La clave no está en hablar más, ni en hablar mejor, sino en asegurar que la comunicación fluya con sentido, representación y justicia, incluso en entornos tecnológicos mediados por algoritmos.

Además de intervenir en los contenidos, la IA está transformando las propias herramientas de comunicación interna, como las intranets corporativas, los portales de empleados o las plataformas colaborativas.

En estos entornos, los algoritmos permiten reorganizar la información, priorizar accesos, recomendar contenidos o generar rutas personalizadas de navegación. Esta reconfiguración técnica puede mejorar la eficiencia y la accesibilidad, pero también introduce riesgos como la pérdida de referencia común, la sobresegmentación funcional o la opacidad en los criterios de personalización.

En este sentido, la integración de IA en canales estructurales debe combinarse con criterios de transparencia, trazabilidad y participación para no erosionar la cohesión interna.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Tabla 9: *Resumen sobre las oportunidades y riesgos de la IA en la comunicación interna*

Aplicación de IA	Beneficios potenciales	Riesgos y alertas
Segmentación de mensajes	Mejora relevancia del contenido Reducción de ruido comunicativo	Exclusión inadvertida Segmentación opaca o no explicada
Análisis de impacto comunicativo	Ajuste de tono y canal a cada perfil Mejora en comprensión	Monitorización invasiva Decisiones automatizadas sin supervisión
Automatización de canales internos	Mayor eficiencia y adaptabilidad en la distribución	No visibilización de grupos sin acceso digital; pérdida de espontaneidad
Personalización de intranets y portales de empleados	Mejora de la experiencia de usuario Acceso inteligente a la información	Desorientación por sobre personalización; pérdida de coherencia narrativa Criterios opacos de visibilidad
Análisis de sentimiento en mensajes	Detección temprana de malestar o desconexión	Lectura emocional malinterpretada Deterioro de la confianza

Fuente: elaboración propia.

1.4.4 Administración de RR. HH.: Eficiencia sí, pero con cuidado

La automatización de tareas administrativas a través de sistemas de IA ha transformado los procesos cotidianos en los departamentos de RR. HH. Aplicaciones como la gestión de turnos, solicitudes de permisos, control horario, actualización de expedientes o procesamiento de nóminas han ganado en rapidez, trazabilidad y reducción de errores humanos (Microsoft WorkLab, 2023; SHRM, 2023). Esta eficiencia operativa representa un avance importante, especialmente en organizaciones de gran escala o con estructuras complejas.

Sin embargo, en este ámbito aparentemente neutro y técnico también se despliegan decisiones que afectan directamente a la equidad, la accesibilidad y el reconocimiento de la diversidad dentro de la plantilla. Automatizar procesos administrativos no puede significar ignorar los matices humanos que estos procesos contienen (CIPD, 2022; Eurofound, 2023).

Uno de los principales riesgos consiste en que los sistemas de IA —por definición programados para optimizar según patrones— no contemplen adecuadamente las excepciones necesarias: licencias por cuidado, adaptaciones para personas con discapacidad, situaciones derivadas de violencia de género o trayectorias laborales no lineales (ILO, 2022). Cuando el estándar es rígido, la automatización puede convertirse en exclusión estructural.

Además de excluir por rigidez normativa, estos sistemas pueden también ser utilizados como instrumentos de control más que de soporte, especialmente cuando el seguimiento de horarios, accesos o rendimiento está automatizado sin mediación humana ni espacios de diálogo. En estos casos, la eficiencia tecnológica puede derivar en mecanismos disciplinarios encubiertos, debilitando la confianza organizativa y erosionando el tejido relacional entre personas y sistemas (AlgorithmWatch, 2022; OECD, 2023).

Además, la transparencia algorítmica y la posibilidad de revisión manual son requisitos fundamentales para garantizar que la administración digital no reproduzca sesgos ni genere barreras invisibles

(Harvard Kennedy School, 2022; IBM, 2023). La falta de supervisión significativa puede llevar a situaciones donde el sistema no solo automatiza tareas, sino también injusticias. Aunque el Reglamento de IA de la UE establece obligaciones específicas para desarrolladores, proveedores y usuarios —precisamente para evitar estos escenarios—, su cumplimiento efectivo sigue siendo un desafío, especialmente cuando las soluciones son desarrolladas o gestionadas por proveedores externos, dificultando la auditoría institucional y la trazabilidad de decisiones.

Por ello, una administración de RR. HH. basada en IA debe combinar eficiencia operativa con sensibilidad institucional. Esto implica:

- Diseñar **sistemas con lógicas flexibles** que contemplen excepciones y trayectorias diversas.
- Establecer **canales claros de retroalimentación y apelación** ante decisiones automatizadas.
- Integrar principios de **equidad, corresponsabilidad y cuidado** en la arquitectura misma de los procesos (World Economic Forum, 2024).

En definitiva, la eficiencia no debe imponerse como una lógica excluyente, sino integrarse en una comprensión amplia y deliberada de lo humano en la administración del trabajo. Esta transformación tecnológica, si se aplica sin contrapesos éticos ni sensibilidad institucional, puede contribuir —como ya se ha observado en algunos sectores— a condiciones laborales más inestables, fragmentadas o sometidas a vigilancia permanente. Solo cuando la IA se orienta a simplificar sin invisibilizar, automatizar sin excluir y modernizar sin deshumanizar, puede cumplir realmente su promesa transformadora.

Tabla 10: *Resumen de las oportunidades y riesgos de las aplicaciones de IA en la administración de Recursos Humanos*

Proceso automatizado	Eficiencia esperada	Riesgos si faltan salvaguardas
Gestión de turnos y horarios	Optimización de recursos, Cobertura ágil de vacantes	Desatención a limitaciones personales Sobrecarga repetida de ciertos perfiles Aumento del riesgo de burnout
Solicitudes y permisos	Rapidez en tramitación Reducción de cargas administrativas	Invisibilización de casuísticas no estándar (cuidado, salud mental, etc.)
Control de accesos y asistencia	Mayor trazabilidad Detección de incidencias	Aplicación disciplinaria sin proporcionalidad ni revisión humana Riesgo de interpretación descontextualizada de los datos
Actualización de expedientes	Centralización eficiente de información	Eliminación de matices relevantes
Procesamiento de nóminas y beneficios	Menor error manual Mayor velocidad de gestión	Fallos sistemáticos no detectados Dificultad de apelación ante errores

Fuente: elaboración propia.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

1.4.5 Prevención de riesgos laborales: Anticipar sin discriminar

La aplicación de IA en el ámbito de la prevención de riesgos laborales (PRL) una oportunidad significativa para avanzar hacia entornos de trabajo más seguros, adaptativos y proactivos. Los sistemas basados en datos permiten identificar patrones de accidentabilidad, detectar zonas de riesgo antes de que se materialicen incidentes y generar recomendaciones preventivas personalizadas para cada perfil o puesto (EU-OSHA, 2023; ILO, 2022). Esta capacidad predictiva puede ayudar a reducir accidentes, optimizar los planes de seguridad y ajustar intervenciones a realidades específicas en tiempo real.

Sin embargo, el paso de una prevención reactiva a una prevención predictiva —aunque tecnológicamente deseable— no está exento de dilemas éticos y organizativos. En particular, existe el riesgo de que las recomendaciones generadas por los modelos deriven en prácticas discriminatorias o excluyentes, sobre todo si no se contextualizan ni se auditan adecuadamente (AlgorithmWatch, 2021; TUC, 2023).

Entre los principales riesgos se encuentran la incorporación de variables sensibles —como el género, la edad, la salud mental o las trayectorias laborales atípicas— como factores de predicción del riesgo. Si no se gestionan con rigor, estas variables pueden derivar en inferencias automatizadas que perpetúan desigualdades estructurales o penalizan determinadas condiciones de salud o estilos de vida. La trazabilidad de estos datos y su uso sin una supervisión humana adecuada puede dificultar el derecho a la explicación, a la rectificación y a la proporcionalidad de las medidas adoptadas.

Uno de los peligros más señalados es el uso de variables sensibles —como el historial médico, el estado emocional percibido o patrones conductuales pasados— como predictores de riesgo, lo cual es especialmente relevante para evitar que personas que han atravesado situaciones de vulnerabilidad —como mujeres víctimas de violencia o personas en procesos de recuperación social— sean penalizadas de forma encubierta. (Eje 1 del proyecto).

Cuando estos datos se interpretan de forma automatizada, sin supervisión humana ni comprensión situacional, pueden dar lugar a la estigmatización de ciertos perfiles, reforzando sesgos previos o atribuyendo “riesgos personales” de forma injustificada (CIPD, 2023; IEEE, 2022).

Además, si los modelos de IA se emplean para anticipar la “idoneidad” de una persona para ocupar o mantener determinados puestos en función de riesgos proyectados, existe la posibilidad de que se generen prácticas de exclusión preventiva: limitar accesos, desviar asignaciones o sobre vigilar a determinados grupos laborales sin valorar adecuadamente los factores estructurales, organizativos o de apoyo que podrían estar implicados (Harvard Business Review, 2022; Eubanks, 2018). Prevenir no equivale a segregar, la anticipación del riesgo debe ir acompañada —cuando sea viable— de medidas de adaptación, acompañamiento y diálogo, no de restricción automatizada o aislamiento sistemático.

Por ello, aplicar IA en prevención no requiere únicamente datos fiables, sino también incorporar principios rectores como la proporcionalidad, la transparencia y la equidad, tal como establece el marco europeo para los sistemas algorítmicos de alto impacto. Aunque estos no forman parte del lenguaje técnico habitual de la prevención de riesgos laborales, son fundamentales para evitar usos automatizados que deriven en exclusión o discriminación. En concreto, se deben cumplir al menos estas condiciones:

• **Auditar** los modelos periódicamente, incluyendo la revisión de variables sensibles y sus impactos diferenciales por género, edad, salud o tipo de contrato (European Commission, 2021).
• Garantizar que la predicción de riesgo **no sustituya el análisis profesional** ni el diálogo con las personas implicadas.
• **Evitar cualquier uso punitivo o excluyente de los modelos**, y asegurar el acceso a mecanismos de revisión o apelación ante decisiones automatizadas.

La integración de estos sistemas debe contemplar también la participación activa de representantes laborales, a fin de garantizar que la prevención tecnológica se alinee con los derechos colectivos y las condiciones reales del trabajo.

La IA puede ser una herramienta valiosa para cuidar mejor a las personas trabajadoras, siempre que se utilice bajo principios de equidad preventiva y no como mecanismo de selección encubierta. En última instancia, la clave está en asegurar que los sistemas anticipan peligros sin crear nuevas formas de vulnerabilidad ni transferir el riesgo a quien ya lo sufre (World Economic Forum, 2024).

La tabla que se presenta a continuación ejemplifica cómo el uso de IA en PRL puede tener un impacto positivo siempre que se oriente a la anticipación y reducción de riesgos, y no se convierta en un instrumento de vigilancia o selección encubierta. La protección de los derechos fundamentales, la participación de las personas trabajadoras y el cumplimiento normativo (como la prohibición del uso de datos biométricos para inferencia emocional establecida por el RIA) deben guiar toda implementación.

Tabla 11: *Usos de la IA en Prevención de Riesgo Laborales*

Aplicación IA en PRL	Potencial beneficio	Riesgos éticos o sociales
Análisis predictivo de accidentes	Anticipar situaciones de riesgo en tiempo real mediante sensores, patrones de comportamiento o historial	Posible uso de datos de siniestralidad con fines no preventivos: limitar acceso a ciertos puestos o atribuir responsabilidades individuales sin análisis contextual
Monitoreo postural con visión por computador	Detectar malas posturas o esfuerzos repetitivos para prevenir trastornos musculoesqueléticos	Riesgo de vigilancia biomecánica constante o penalización por estilos de trabajo divergentes, sin atención a diversidad funcional o contextos individuales
Modelos de fatiga basados en IA	Prever situaciones de sobrecarga física o mental para ajustar turnos o tareas	Inferencia emocional no consentida o basada en datos biométricos (prohibido por el RIA)
Sistemas de mantenimiento predictivo	Prevenir fallos técnicos que puedan generar riesgos en el entorno laboral	Posible desplazamiento de responsabilidades a trabajadores si no se interpreta bien la causa del error
IA en evaluaciones de riesgo psicosocial	Identificar factores de estrés organizacional o dinámicas de acoso mediante análisis de lenguaje	Sesgos en la interpretación algorítmica del lenguaje; ausencia de validación humana y participación de los actores implicados
Recomendaciones personalizadas de medidas preventivas	Ajustar la formación o las medidas a las características del puesto o del trabajador	Riegos de sobre individualización, sin abordar causas estructurales o colectivas de los riesgos

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Tras exponer los riesgos éticos potenciales y los dilemas asociados al uso de IA en prevención de riesgos laborales, resulta pertinente mostrar cómo estas dinámicas se concretan en la práctica. La siguiente tabla recoge ejemplos reales de aplicación de sistemas de IA en entornos laborales, ilustrando tanto su utilidad preventiva como los retos que plantean.

Tabla 12: *Aplicaciones de IA en PRL*

Aplicación de IA en PRL	Beneficios Potenciales	Riesgos Éticos o Sociales
Detección de uso inadecuado de EPIs	Identificar en tiempo real la ausencia o mal uso de equipos de protección individual	Posible percepción de vigilancia excesiva o sanción sin contexto adecuado
Análisis de posturas y movimientos	Prevenir trastornos musculoesqueléticos mediante la detección de posturas forzadas	Riesgo de interpretar incorrectamente comportamientos individuales sin considerar factores contextuales
Monitoreo de fatiga y estrés	Anticipar situaciones de sobrecarga física o mental para ajustar turnos o tareas	Inferencia emocional no consentida o basada en datos biométricos, lo cual está prohibido por el RIA
Detección de zonas peligrosas y acceso no autorizado	Alertar sobre la presencia en áreas de riesgo o restringidas en tiempo real	Posible invasión de la privacidad y generación de desconfianza entre los trabajadores
Análisis predictivo de accidentes	Identificar patrones que podrían indicar una mayor probabilidad de accidentes	Discriminación algorítmica hacia trabajadores con mayor historial o que desempeñan tareas más expuestas
Automatización de tareas Repetitivas y peligrosas	Reducir la exposición a riesgos y permitir que los trabajadores se enfoquen en tareas de mayor valor añadido	Desplazamiento de responsabilidades a trabajadores si no se interpreta bien la causa del error
Formación personalizada en PRL	Adaptar la formación según las características y necesidades específicas de cada trabajador.	Riesgo de sobre individualización, sin abordar causas estructurales o colectivas de los riesgos
Simulación de situaciones de riesgo mediante realidad virtual	Sensibilizar sobre los riesgos laborales y sus consecuencias de manera inmersiva	Posible falta de accesibilidad para todos los trabajadores o dependencia excesiva de la tecnología

Fuente: Asepeyo.

Estos casos prácticos muestran que, cuando la IA se alinea con principios como la equidad preventiva, puede transformar la PRL sin generar nuevas formas de vulnerabilidad.

1.4.6 Relaciones laborales: hacia una negociación algorítmica responsable

La IA está empezando a introducirse en un terreno particularmente sensible de las organizaciones: las relaciones laborales. Lejos de limitarse a tareas técnicas, su aplicación se extiende a procesos que tocan directamente el equilibrio entre las partes, como el análisis automatizado de convenios colectivos, la simulación de escenarios de negociación, o la detección temprana de tensiones laborales. En estos espacios, la IA puede aportar mayor agilidad documental, precisión en el tratamiento normativo y capacidad prospectiva para planificar conflictos o acuerdos (ILO, 2023; SHRM, 2023).

A este potencial técnico se suma un primer riesgo crítico: la gestión de la información laboral en entornos no seguros o sin validación legal adecuada. En particular, el uso de sistemas de IA generativa abiertos —como asistentes conversacionales o plataformas de respuesta automática— para interpretar normativas, convenios o derechos laborales puede conducir a decisiones erróneas o mal fundamentadas. Estos modelos, aunque útiles en tareas generales, carecen de la precisión contextual necesaria para abordar la complejidad de las relaciones laborales, donde entran en juego no solo leyes y convenios colectivos, sino también jurisprudencia, acuerdos internos y casuística sectorial. Su uso sin supervisión jurídica puede traducirse en incumplimientos involuntarios, conflictos innecesarios o vulneración de derechos adquiridos.

En este sentido, la aplicación de IA en el ámbito laboral debe distinguir claramente entre herramientas de apoyo documental o simulación, y aquellas que influyen en decisiones estratégicas o negociadoras. Solo las primeras pueden utilizarse con márgenes de error aceptables; las segundas requieren validación humana, trazabilidad jurídica y participación de las partes interesadas.

Sin embargo, esta capacidad tecnológica conlleva riesgos específicos cuando se emplea sin transparencia ni control institucional, especialmente en entornos donde el derecho a la representación colectiva, la negociación sindical y el ejercicio de libertades fundamentales se encuentra en juego (OECD, 2022; European Commission, 2021).

Uno de los principales riesgos es el uso de datos históricos o comportamientos sindicales agregados para anticipar estrategias de negociación o diseñar respuestas automatizadas frente a dinámicas de conflicto. Este tipo de analítica puede derivar en prácticas antisindicales encubiertas si se utiliza para debilitar la capacidad organizativa de los representantes de los trabajadores, evitar paros o neutralizar la presión colectiva, incluso sin intención explícita (AlgorithmWatch, 2021; ETUI, 2022). En estos casos, la IA se transforma en una herramienta de gestión del riesgo sindical, lo que puede vulnerar directamente el derecho a la libertad sindical (TUC, 2023).

Otro ámbito crítico es la introducción de sistemas de IA en procesos de mediación, arbitraje o resolución de disputas laborales. Aunque estas herramientas pueden ofrecer modelos de simulación y facilitar el acceso a precedentes normativos, su uso debe estar acompañado de supervisión humana significativa y trazabilidad completa de las decisiones (Harvard Law School, 2022; European Commission, 2021). Automatizar sin transparencia en este tipo de contextos puede deteriorar la confianza entre las partes, deslegitimar los acuerdos o reforzar desequilibrios de poder existentes.

La aplicación de IA en las relaciones laborales no debe orientarse a reducir el conflicto en sí mismo, sino a transformarlo en procesos de diálogo más informado, más justo y eficaz.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Para ello, se requieren:

• Mecanismos de **participación sindical** en la implementación de sistemas de IA que afecten la negociación colectiva (World Economic Forum, 2023).

• Garantías de que los datos analizados **no serán utilizados para vigilancia política o discriminación ideológica.**

• **Marcos de gobernanza** que aseguren que la IA fortalece el derecho laboral, no lo debilita (Eurofound, 2023).

• **Limitación** del uso de modelos generativos generalistas en procesos de toma de decisiones laborales, especialmente cuando implican interpretación normativa o aplicación de convenios colectivos.

Buenas prácticas para una negociación colectiva asistida por IA

• Garantizar la participación sindical efectiva en el diseño, prueba y despliegue de los sistemas algorítmicos que impacten en la negociación.
• Establecer protocolos de auditoría y trazabilidad sobre el uso de datos sindicales o de comportamiento colectivo.
• Asegurar que la inteligencia artificial se utiliza para informar y facilitar el diálogo, no para debilitarlo ni predecir respuestas sindicales.
• Prohibir expresamente el uso de sistemas algorítmicos para fines de perfilado ideológico o disuasión del derecho de huelga.
• Incorporar principios de proporcionalidad, consentimiento y no discriminación en todos los sistemas aplicados a la mediación laboral.

En última instancia, el reto no es solo técnico ni jurídico, sino profundamente democrático: asegurar que los algoritmos no desdibujan el papel de los actores sociales, sino que lo amplifican en condiciones de justicia, trazabilidad y control institucional.

1.5 Impacto de la IA en materia de diversidad e inclusión

Si bien en el apartado de Gobernanza del dato se abordó la importancia de que esta sea robusta como base para una IA inclusiva, en este capítulo analizamos cómo se manifiestan esos desafíos en la práctica y qué efectos concretos tienen sobre la diversidad y la inclusión en el entorno laboral. Nos centramos en colectivos específicos, en estrategias de mitigación organizativa y en el papel clave de la sensibilización, formación y participación.

La IA aplicada al ámbito laboral no es neutral. Su impacto sobre la diversidad y la inclusión depende de múltiples factores: desde la calidad de los datos y el diseño algorítmico, hasta la cultura organizativa y las prácticas de comunicación.

Mientras que el gobierno del dato robusto —como se analizó en el apartado referido a la gobernanza— establece las condiciones técnicas y normativas para mitigar el sesgo, este apartado examina los efectos prácticos y sociales que la IA puede tener sobre colectivos históricamente infrarrepresentados, así como las estrategias necesarias para evitar que la automatización profundice las desigualdades existentes.

La adopción acrítica de sistemas de IA en la gestión de personas puede reproducir estereotipos, invisibilizar trayectorias no normativas o excluir a individuos por no encajar en patrones dominantes. Pero también puede ser una herramienta para ampliar la inclusión, eliminar sesgos humanos y construir organizaciones más diversas y justas. El resultado depende, en gran medida, de cómo se diseñe, implemente y supervise la IA en cada contexto.

1.5.1 Infrarrepresentación de colectivos desfavorecidos

La infrarrepresentación algorítmica de colectivos vulnerables es una de las vías más silenciosas pero persistentes de reproducción de desigualdades en el entorno laboral. Cuando los datos históricos excluyen trayectorias atípicas, el resultado es un sistema que automatiza la invisibilidad.

En el contexto laboral, la IA tiene el potencial de transformar los procesos de gestión de personas, pero también de consolidar desigualdades estructurales si no se aplican salvaguardas adecuadas. Una de las manifestaciones más significativas de este riesgo es la infrarrepresentación de colectivos históricamente desfavorecidos en los sistemas automatizados de toma de decisiones (Buolamwini & Gebru, 2018; Raghavan et al., 2019).

La causa principal suele ser la utilización de datos históricos sesgados, que reflejan realidades pasadas marcadas por la desigualdad. Estos datos, al ser utilizados sin un análisis crítico, pueden introducir distorsiones en los modelos predictivos, penalizando a personas que no encajan en los perfiles "típicos" identificados por los algoritmos. Por ejemplo, mujeres con trayectorias laborales interrumpidas, personas mayores sin experiencia digital, o candidatos de origen migrante con titulaciones no homologadas pueden ser evaluados de forma menos favorable, no por su competencia real, sino por la falta de representatividad de sus perfiles en los conjuntos de entrenamiento (Ajunwa, 2020; European Commission, 2020).

Además, debe considerarse que el propio marco de protección de datos, fundamental para salvaguardar los derechos individuales, establece restricciones sobre el tratamiento de categorías especiales de información, como la situación de discapacidad, la condición de víctima de violencia de género, antecedentes penales o enfermedades crónicas. Esta limitación, aunque necesaria para preservar la privacidad, dificulta la inclusión explícita de estos colectivos en los conjuntos de entrenamiento, perpetuando en algunos casos su invisibilidad estadística y limitando la capacidad de los sistemas para corregir sesgos estructurales (European Data Protection Board, 2021).

Esta infrarrepresentación no es solo un fallo técnico: es una cuestión de justicia social. Los algoritmos, además de clasificar, también asignan oportunidades, y cuando lo hacen basándose en patrones excluyentes, contribuyen a invisibilizar a quienes ya enfrentan barreras sistémicas. En este sentido, la IA no reproduce únicamente sesgos de género, edad, discapacidad u origen étnico por separado, sino que puede amplificarlos cuando se combinan.

Esta interseccionalidad de los sesgos —el efecto acumulado de múltiples factores de exclusión— debe ser tenida en cuenta en toda auditoría de sistemas algorítmicos (Barocas, Hardt & Narayanan, 2019; Binns, 2018).

a) Representación crítica y validación diferenciada

Para mitigar estos riesgos, es esencial adoptar una aproximación deliberada a la representatividad. Esto implica revisar los datos de entrenamiento desde una mirada crítica, que no solo evalúe la calidad y completitud técnica, sino también la equidad distributiva de los perfiles incluidos. Asimismo, deben desarrollarse mecanismos de validación diferenciada, que permitan evaluar el rendimiento de los algoritmos por subgrupos poblacionales y no solo en términos globales (Holstein et al., 2019).

Una práctica emergente es la creación de conjuntos de validación específicos por colectivo, que

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

permitan detectar sesgos asimétricos en las recomendaciones o decisiones automatizadas. Estas pruebas son útiles para el ajuste técnico de los modelos y, además, representan una herramienta de transparencia hacia la organización y los colectivos afectados (Obermeyer et al., 2019).

b) Inclusión activa en el diseño de datos

Más allá de la revisión crítica, es necesario fomentar una inclusión activa de colectivos subrepresentados en las etapas de diseño y recopilación de datos. Esto puede lograrse mediante el uso de datos complementarios, la generación de ejemplos sintéticos bien calibrados o la incorporación de voces diversas en la definición de las variables relevantes (Eubanks, 2018; Noble, 2018). En lugar de corregir a posteriori los sesgos de los datos, esta perspectiva busca prevenir la exclusión desde el origen del sistema.

En este sentido, abordar la infrarrepresentación de colectivos desfavorecidos no solo responde a un principio de equidad organizativa, sino que también actúa como palanca para promover la autonomía personal y la inclusión social (Eje 1) de personas en situación de vulnerabilidad, como migrantes, personas con discapacidad o víctimas de violencia de género. Además, garantizar la representatividad en los datos y sistemas de IA mejora directamente la empleabilidad de estos colectivos (Eje 5), permitiendo procesos de selección más objetivos y trayectorias laborales más estables.

1.5.2 Papel de la sensibilización y la comunicación

Explicar la IA no es una cuestión técnica: es una cuestión de poder. Lo que no se comunica, no se cuestiona; y lo que no se cuestiona, se normaliza.

Los sistemas de IA no actúan de forma aislada: están insertos en contextos organizativos que moldean su uso, su comprensión y su legitimidad. La sensibilización en torno al uso de sistemas de IA no solo favorece una cultura organizativa más crítica y participativa, sino que constituye una herramienta central para promover los derechos digitales y reducir la brecha digital (Eje 3). Al facilitar el acceso comprensible a la información sobre los algoritmos y su impacto, especialmente entre la población más joven y los colectivos vulnerables, se refuerzan los principios de inclusión digital y democratización tecnológica. Una herramienta clave en este ámbito son los portales de explicabilidad, plataformas digitales que permiten conocer de forma clara y accesible los criterios de decisión algorítmica (European Commission, 2020). Estas acciones permiten que todos los empleados, independientemente de su nivel de formación o condición social, puedan ejercer un control real sobre las decisiones automatizadas que les afectan.

Por ello, la forma en que se comunica la IA dentro de las organizaciones influye directamente en su aceptación, supervisión y capacidad transformadora. La presentación de estos sistemas como neutrales, objetivos o “puramente técnicos” puede generar una falsa confianza, cuando en realidad traducen valores, decisiones y supuestos codificados en su diseño (Ajunwa, 2020; Dankloff et al., 2024).

En este sentido, la sensibilización no es un complemento, sino un eje central de toda estrategia inclusiva basada en inteligencia artificial. La automatización sin comunicación crítica puede deshumanizar procesos y reforzar la idea de que los algoritmos son incuestionables, incluso cuando producen decisiones que afectan profundamente a la trayectoria laboral de las personas (OECD, 2025; Mojsilovic, 2023).

a) De la alfabetización técnica a la narrativa responsable

Una cultura organizativa que aspire a ser diversa y justa debe articular una narrativa clara, comprensible y honesta sobre el uso de IA. Esto implica explicar cómo funciona un sistema y también por qué se ha implementado, qué decisiones automatiza, qué riesgos conlleva y cómo se pueden revisar sus resultados (Ehsan et al., 2021; Holstein et al., 2019). Esta comunicación debe estar adaptada a los diferentes niveles de la organización. No es lo mismo explicar un sistema algorítmico a personal técnico que a representantes sindicales o a personas que forman parte de colectivos con menor familiaridad tecnológica. En estos casos, es esencial emplear un lenguaje accesible y emocionalmente relevante, capaz de generar confianza sin caer en el tecnicismo ni en el marketing (European Commission, 2020; Binns, 2018).

La alfabetización crítica en IA emerge como una práctica fundamental. A diferencia de la formación meramente instrumental, la alfabetización crítica permite identificar los sesgos que pueden contener los sistemas, entender sus límites y ejercer los derechos digitales de forma informada (Barocas, Hardt & Narayanan, 2019). Las campañas internas deben ir más allá de lo informativo para ser también transformadoras: no se trata solo de conocer la IA, sino de relacionarse críticamente con ella.

Un error frecuente en las estrategias de comunicación sobre inteligencia artificial es enfocarlas exclusivamente en el desarrollo de habilidades técnicas, competencias STEM o capacidades de uso de herramientas específicas. Este enfoque, aunque necesario, resulta insuficiente para acompañar de manera justa y equitativa la implementación y despliegue de la inteligencia artificial en las organizaciones.

Una comunicación transformadora debe ir más allá: tiene que incorporar información clara sobre el impacto de los sesgos algorítmicos, las implicaciones de la legislación de protección de datos, la necesidad de actuar con pensamiento crítico y los retos que plantea la irrupción generalizada de la IA en todos los ámbitos de la sociedad.

Este planteamiento estratégico, además, no debe limitarse al ámbito organizativo: es necesario que sea concebido como un esfuerzo escalable que, a nivel país, requiere el liderazgo activo de los poderes públicos para garantizar una alfabetización crítica inclusiva y accesible (, 2021).

b) Comunicación como derecho y herramienta de justicia

Además de promover la comprensión, la comunicación organizativa debe también garantizar la posibilidad de cuestionar, de expresar preocupaciones y de participar activamente en la mejora de los sistemas algorítmicos. Esto supone habilitar canales donde las personas puedan pedir explicaciones sobre decisiones automatizadas, corregir errores detectados o incluso proponer mejoras (Eubanks, 2018; IBM, 2023). La transparencia sin participación se convierte en un ejercicio simbólico; solo cuando existe interacción efectiva, la comunicación puede contribuir a una inteligencia artificial realmente inclusiva.

Esta dimensión participativa de la comunicación resulta especialmente crucial para los colectivos en situación de vulnerabilidad. Cuando las personas sienten que no entienden el sistema, que no pueden influir en él o que sus dudas no encuentran respuesta, se produce un desplazamiento del control y un debilitamiento de la confianza institucional. En cambio, una comunicación inclusiva y deliberativa refuerza el vínculo entre innovación tecnológica y justicia organizativa (Dankloff et al., 2024).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

1.5.3 Formación [4]

Sin formación, no hay justicia algorítmica posible. Una IA ética requiere no solo buenos datos, sino personas capaces de comprenderla, cuestionarla y gobernarla.

Además de tecnologías responsables, el despliegue de sistemas de IA en el ámbito laboral requiere personas preparadas para comprender, supervisar y orientar el uso de esos sistemas. En este contexto, la formación adquiere un papel central para prevenir exclusión, reducir desigualdades y construir organizaciones más justas (European Commission, 2025; Trustible, 2025).

No se trata únicamente de impartir contenidos técnicos a desarrolladores. Lo que se necesita es un ecosistema formativo transversal, que involucre a toda la organización desde su rol específico y con materiales adaptados a su nivel de responsabilidad, comprensión y decisión (AIHR, 2025a; IBM, 2025).

a) Formación diferenciada por públicos clave

Tabla 13: *Públicos estratégicos y contenidos de formación que deben abordarse para garantizar un uso ético, inclusivo y transparente de la IA*

Público	Contenidos clave
Técnicos de datos y desarrolladores	Identificación y mitigación de sesgos algorítmicos, validación interseccional, aplicación de fairness metrics, impacto social de las decisiones técnicas (AI CERTs, 2025)
Equipos de RR. HH.	Interpretación de salidas algorítmicas, riesgos en selección automatizada, rediseño de entrevistas con apoyo IA, sesgos en evaluación del desempeño (AIHR, 2025b; LinkedIn Learning, 2025)
Mandos intermedios y dirección	Comunicación de decisiones basadas en IA, implicaciones éticas del uso organizativo de algoritmos, rendición de cuentas y supervisión humana significativa (Stanmore School, 2025)
Toda la plantilla	Derechos digitales, funcionamiento básico de los sistemas algorítmicos, canales de reclamación y mecanismos de revisión de decisiones (Coursera, 2025a)
Intermediadores laborales, profesionales sociales, RLT y agentes públicos de empleo	Detección de sesgos algorítmicos en procesos de acceso al empleo, defensa de los derechos digitales laborales, mecanismos de reclamación y supervisión de la equidad en la aplicación de la IA (OIT, 2021)

Fuente: elaboración propia.

Segmentar los contenidos de la formación por públicos estratégicos garantiza y contribuye a garantizar un uso ético, inclusivo y transparente de la IA, como apunta la tabla. Esta segmentación

[4] **Nota**: Olga Herrero es miembro del Consejo Asesor del Proyecto IA+Igual y ha aportado su experiencia en este epígrafe como experta en Formación empresarial y e-learning.

permite que cada grupo, ya sea parte de la organización o actuando como intermediador social, pueda interiorizar los riesgos y potencialidades de la IA desde su posición específica, fomentando una cultura organizativa compartida, pero también un ecosistema laboral más justo, transparente e inclusivo.

Un ecosistema formativo inclusivo también contribuye directamente a mejorar la empleabilidad de las personas en situación de vulnerabilidad (Eje 5), proporcionando competencias críticas para un entorno laboral cada vez más automatizado. Diseñar itinerarios formativos adaptados a los perfiles técnicos y a aquellos colectivos tradicionalmente alejados del ámbito digital favorece su inserción sociolaboral y reduce las barreras de acceso a oportunidades profesionales sostenibles. Asimismo, incorporar metodologías accesibles y gamificadas refuerza la autonomía personal y la inclusión social (Eje 1), especialmente en grupos jóvenes o en riesgo de exclusión.

b) Formación viva y empática

El impacto real de estas formaciones no depende únicamente de su contenido, sino también de su metodología. Las acciones más eficaces son aquellas que se basan en:

- **Casos reales,** que muestran situaciones concretas en las que la IA ha producido decisiones erróneas, injustas o inesperadas.
- **Simulaciones interactivas,** que permiten a los participantes experimentar los procesos algorítmicos desde múltiples perspectivas (Coursera, 2024).
- **Espacios deliberativos,** donde puedan expresarse dudas, preocupaciones o propuestas de mejora.

Este enfoque dinámico favorece la empatía y la comprensión transversal. Lejos de convertir la formación en un trámite, la transforma en un instrumento de cambio cultural, donde cada persona se reconoce como agente activo en la construcción de una IA más justa. Además, este enfoque formativo no solo sensibiliza, sino que fortalece estructuras de gobernanza responsable al distribuir conocimiento clave en toda la cadena de decisión. Este tipo de metodologías también contribuyen a reducir la ansiedad tecnológica, proporcionando un entorno de aprendizaje seguro y accesible para todos los perfiles (Tarafdar, Cooper & Stich, 2019).

c) Principales áreas de impacto

La incorporación de IA en los procesos de gestión de personas no es una simple cuestión tecnológica. Su efecto en la diversidad y la inclusión laboral depende de cómo, cuándo y con qué objetivos se despliega. Si se integra de manera crítica, ética y supervisada, puede actuar como acelerador de justicia organizativa (HealthboxHR, 2025; LinkedIn, 2025). Por el contrario, un uso acrítico, automatizado o descontextualizado puede profundizar desigualdades existentes y generar nuevas formas de exclusión (Wired, 2025).

Este subapartado identifica las principales áreas donde el impacto de la IA se hace especialmente visible, tanto en su dimensión de riesgo como en su potencial transformador. Se trata de áreas que concentran decisiones clave sobre trayectorias profesionales, oportunidades de acceso, reconocimiento y representación (Bano et al., 2024).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

d) Áreas de impacto estratégicas y bifurcación de consecuencias

En cada una de estas áreas de impacto, la IA actúa como un multiplicador: puede amplificar sesgos, pero también oportunidades. La diferencia reside en el diseño del sistema, la calidad de los datos, el control humano y la participación de los colectivos afectados (Plum, 2025; The Times, 2024).

Tabla 14: *Potencial y riesgo de la IA según como se aplique en cinco áreas de impacto de RR. HH.*

Área de impacto	Riesgo si se aplica mal	Potencial si se aplica bien
Reclutamiento	Exclusión de perfiles atípicos o invisibilizados	Acceso justo y ampliado mediante matching inclusivo
Evaluación de desempeño	Penalización de trayectorias no normativas	Recompensa basada en datos reales, no en valoraciones inadecuadas
Promoción y carrera	Repetición de sesgos históricos	Identificación proactiva de talento diverso
Clima laboral	Desconfianza en decisiones automatizadas	Fortalecimiento del compromiso si hay transparencia
Representación interna	Marginalización de voces distintas	Visibilización y empoderamiento de colectivos subrepresentados

Fuente: elaboración propia.

En estas cinco áreas de impacto, la IA decide más que recomienda. Gestionarlas con responsabilidad es una forma directa de proteger la justicia organizativa y de garantizar que la automatización no erosione, sino que amplíe, los márgenes de equidad dentro de la empresa.

e) Del riesgo a la oportunidad: elementos de inflexión

Los riesgos no son inevitables. Existen elementos de inflexión que permiten orientar el impacto de la IA hacia modelos más inclusivos. Algunos de ellos son:

- **Transparencia activa:** cuando las reglas de decisión algorítmica son conocidas, es más fácil detectar sesgos y corregir errores (Qualee, 2025).
- **Participación plural:** la inclusión de distintos grupos en la definición y validación de sistemas permite evitar la reproducción de sesgos invisibles (Macorva, 2024).
- **Supervisión humana significativa:** las decisiones más críticas no deben ser delegadas completamente a la herramienta. El juicio humano debe mantenerse como filtro deliberativo (People Matters Global, 2025).
- **Ajuste contextualizado:** los modelos deben adaptarse a la realidad y cultura de la organización, no a estándares abstractos o importados (Forbes, 2023).

Transformar las áreas de riesgo en palancas de cambio no es solo una posibilidad técnica. Es una elección institucional que requiere visión estratégica, compromiso ético y gobernanza participativa.

La identificación de estas áreas de impacto permite gestionar los riesgos técnicos de la IA y, además, ofrece una hoja de ruta para alinear la transformación digital con los derechos sociales y laborales.

Cuando se incorpora la IA con criterios de inclusión y ética, se favorece tanto la inserción sociolaboral de colectivos vulnerables (Eje 5) como la autonomía de las personas en su recorrido profesional (Eje 1), al evitar que trayectorias no normativas sean penalizadas por patrones históricos. A su vez, garantizar una implementación explicable y participativa contribuye a reducir la brecha digital (Eje 3), asegurando que los avances tecnológicos no queden reservados a una élite tecnificada, sino que beneficien a toda la diversidad de la fuerza laboral.

1.5.4 Principales áreas de impacto

A lo largo de este apartado se ha puesto de manifiesto que la IA no es solo una herramienta técnica, sino también un factor estructural en la distribución de oportunidades laborales. Su capacidad para ampliar la inclusión o profundizar la desigualdad depende en gran medida del modo en que se diseña, se comunica, se implementa y se supervisa.

Hemos identificado cuatro palancas clave para transformar el impacto de la IA en el ámbito laboral:

1. Reconocer la infrarrepresentación estructural de ciertos colectivos en los datos y modelos, y actuar desde una lógica reparadora e interseccional.

2. Desplegar estrategias sólidas de sensibilización y comunicación, que permitan comprender los sistemas algorítmicos y participar activamente en su mejora.

3. Construir un ecosistema formativo transversal y contextualizado, que prepare a todos los niveles de la organización para una IA ética y humanamente comprensible.

Las cuatro palancas para una inteligencia artificial laboral inclusiva

- Reconocer la infrarrepresentación estructural, y actuar desde una lógica reparadora e interseccional.
- Sensibilizar y comunicar con transparencia, para empoderar y permitir participación.
- Formar a toda la organización, asegurando comprensión, cuestionamiento y gobernanza compartida.
- Gestionar estratégicamente los ejes de impacto, convirtiendo riesgos en oportunidades para la equidad.

4. Identificar y gestionar los ejes críticos de impacto, convirtiendo los riesgos en oportunidades a través de transparencia, supervisión humana y participación plural.

Este enfoque requiere algo más que buenas intenciones: necesita estructuras, procesos y recursos. Pero, sobre todo, demanda una visión institucional clara sobre el tipo de organización que se quiere construir en la era algorítmica.

Integrar la diversidad y la inclusión en el ciclo de vida de la IA no es solo una obligación normativa ni un imperativo ético: es una condición de posibilidad para que la transformación digital sea también una transformación democrática y socialmente sostenible.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

2. MARCO LEGAL PARA IMPLANTAR LA IA EN EL ÁMBITO LABORAL

La implementación de inteligencia artificial en recursos humanos (RR. HH.) plantea desafíos legales, éticos y operativos significativos, especialmente en la protección de datos, la prevención de sesgos, la garantía de transparencia y la supervisión humana. Esta revolución tecnológica, aunque ofrece oportunidades notables en términos de eficiencia, personalización y toma de decisiones basada en datos, plantea también desafíos legales, éticos y sociales de primer orden. La creciente adopción de sistemas automatizados en RR. HH. exige una respuesta normativa rigurosa, centrada en la protección de los derechos fundamentales de las personas, la transparencia en el uso de algoritmos y la garantía de procesos justos y auditables. En este contexto, el marco legal vigente en Europa y España —con especial relevancia del Reglamento General de Protección de Datos (RGPD), la Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales (LOPDGDD) y la Carta de Derechos Digitales— establece obligaciones estrictas sobre el consentimiento informado, la no discriminación y la supervisión humana en la toma de decisiones automatizadas.

Este documento parte de la premisa de que la integración de la IA en RR. HH. debe regirse por líneas estratégicas que garanticen la inclusión, la equidad y la protección de los colectivos más vulnerables. Así, se integran los tres ejes fundamentales de este proyecto: la promoción de los derechos de las personas, la transformación digital inclusiva y la mejora de la inserción sociolaboral. Estos ejes están alineados con políticas públicas y directrices europeas sobre supervisión algorítmica, así como con los Objetivos de Desarrollo Sostenible (ODS) relativos a la igualdad de género, el trabajo decente y la reducción de las desigualdades.

A lo largo del documento se examinan los fundamentos legales, las líneas estratégicas y los retos emergentes, con especial atención a la prevención de sesgos, la equidad algorítmica y la protección de los derechos fundamentales. Se ofrecen, además, orientaciones metodológicas y recomendaciones para garantizar que la IA, lejos de perpetuar desigualdades, contribuya a una gestión del talento más justa, inclusiva y socialmente responsable

2.1 Marco legal en Europa [5]

2.1.1 Reglamento General de Protección de Datos (RGPD)

El Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (RGPD) es uno de los pilares fundamentales para el uso de IA en la gestión de personas, regulando el tratamiento de datos personales en procesos como la selección de candidatos, evaluación del desempeño, gestión del talento y prevención de riesgos laborales.

[5] **Nota**: El autor de los apartados 2.1 Marco legal en europa, 2.2 Marco legal en España y 2.3 Marco legal global e internacional de este capítulo es Efrén Díaz Díaz, abogado, doctor en Derecho por la Universidad de Navarra y responsable de las Áreas de Tecnología y Derecho Espacial del Bufete Mas y Calvet.

a) Consentimiento informado

En el contexto de RR. HH., los candidatos deben otorgar un consentimiento explícito y específico (artículo 6.1.a, Reglamento (UE) 2016/679) para el uso de sistemas de IA, como algoritmos de filtrado de currículums, análisis de entrevistas mediante reconocimiento facial o evaluación de desempeño basada en datos biométricos.

Por ejemplo, una empresa que utilice IA para analizar el tono de voz en entrevistas debe informar a los candidatos sobre el propósito, el alcance y los riesgos del tratamiento, obteniendo su consentimiento previo.

La falta de consentimiento puede derivar en sanciones significativas, con multas de hasta 20 millones de euros o el 4 % de la facturación anual global (Artículo 83, RGPD).

b) Transparencia y derecho de información

El artículo 13 del Reglamento (UE) 2016/679 establece el deber de información:

> *"Cuando se recojan datos personales relativos a un interesado, el responsable del tratamiento, en el momento en que se obtengan dichos datos, le facilitará toda la información indicada".*

Esto obliga a las empresas a detallar cómo funcionan los sistemas de IA, incluyendo los criterios algorítmicos y los datos utilizados. Por ejemplo, un sistema de inteligencia artificial que priorice candidatos con experiencia en ciertas industrias debe informar a los solicitantes sobre los parámetros de selección. La falta de transparencia puede violar el derecho de información y derivar en reclamaciones legales.

c) Prevención de sesgos y no discriminación

El artículo 5.1.f del Reglamento (UE) 2016/679 se destaca este importante principio relativos al tratamiento:

> *"Los datos personales serán tratados de manera que se garantice una seguridad adecuada de los datos personales, incluida la protección contra el tratamiento no autorizado o ilícito".*

Los algoritmos de IA deben ser diseñados y auditados para evitar sesgos discriminatorios basados en género, raza, edad, orientación sexual u otras características protegidas.

d) Derecho a no ser objeto de decisiones automatizadas

En materia de decisiones individuales automatizadas, incluida la elaboración de perfiles, el artículo 22 del RGPD establece un derecho con particular incidencia práctica:

> *"El interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar".*

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

En RR. HH., esto implica que decisiones como contrataciones, promociones o despidos no pueden basarse exclusivamente en algoritmos. Por ejemplo, un sistema de IA que evalúe el desempeño de un empleado mediante métricas automatizadas debe ser revisado por un supervisor humano antes de tomar decisiones que afecten su empleo. Este derecho protege la autonomía y dignidad de los trabajadores, evitando la deshumanización de los procesos laborales.

e) Evaluaciones de impacto en protección de datos (EIPD)

> *"Cuando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento realizará, antes del tratamiento, una evaluación del impacto de las operaciones de tratamiento"* (Artículo 35 RGPD).

Las empresas que utilicen IA en RR. HH. deben realizar EIPD para identificar riesgos, como la discriminación algorítmica o la vulneración de la privacidad, y proponer medidas de mitigación. Por ejemplo, una empresa que implemente un sistema de supervisión del desempeño basado en IA tiene que evaluar los riesgos de vigilancia excesiva.

2.1.2 Reglamento de Inteligencia Artificial europeo (RIA)

La Unión Europea lidera la regulación de la IA con el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial). Es conocido como AI Act, entró en vigor el 1 de agosto de 2024 y será completamente aplicable en 2026.

Este reglamento establece un marco basado en el riesgo para los sistemas de IA, con implicaciones específicas en el ámbito laboral.

a) Clasificación de riesgos en el AI Act [6]

El AI Act categoriza cuatro niveles de riesgo: inaceptable, alto, limitado y mínimo.

Los sistemas de RR. HH., como los utilizados en contratación, evaluación del desempeño, asignación de tareas o prevención de riesgos laborales, suelen clasificarse como de alto riesgo:

> *"Los sistemas de inteligencia artificial de alto riesgo deberán cumplir con requisitos específicos, incluyendo la gestión de riesgos, la calidad de los datos, la transparencia, la supervisión humana, la documentación, la trazabilidad y la ciberseguridad"* (Artículos 8-15, Reglamento (UE) 2024/1689).

b) Requisitos para sistemas de alto riesgo:

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de IA).

[6] **Nota**: Como se ha explicado en la Introducción del Libro Blanco, en el documento se utilizan indistintamente algunos términos en castellano e inglés. Por ejemplo: Reglamento de IA (RIA) y AI Act. En este caso, se usa también la denominación Reglamento (UE) 2024/1689.

• Gestión de riesgos:

> *"Los proveedores de sistemas de IA de alto riesgo establecerán un sistema de gestión de riesgos que identifique, evalúe y mitigue los riesgos durante todo el ciclo de vida del sistema"* (Artículo 9, Reglamento (UE) 2024/1689).

Las empresas deben identificar riesgos como sesgos algorítmicos que puedan discriminar a ciertos grupos. Por ejemplo, un sistema de selección que favorezca a candidatos jóvenes debe ser ajustado para garantizar la equidad.

• Datos de alta calidad:

> *"Los conjuntos de datos utilizados para el entrenamiento, validación y prueba de sistemas de IA de alto riesgo deberán ser pertinentes, representativos, sin errores y completos"* (Artículo 10, Reglamento (UE) 2024/1689).

Esto implica utilizar datos diversos y libres de sesgos históricos. Por ejemplo, si los datos de entrenamiento de un algoritmo de selección provienen de una empresa con poca diversidad de género, el algoritmo puede perpetuar esa falta de representación.

• Transparencia:

> *"Los sistemas de IA de alto riesgo deberán diseñarse y desarrollarse de manera que permitan a los usuarios comprender y utilizar adecuadamente el sistema"* (Artículo 13, Reglamento (UE) 2024/1689).

Los empleados y candidatos deben ser informados sobre cómo funciona el sistema de IA, sus limitaciones y los derechos asociados, como el derecho a revisión humana.

• Supervisión humana:

> *"Los sistemas de IA de alto riesgo deberán diseñarse y desarrollarse de manera que puedan ser supervisados eficazmente por personas físicas durante su uso"* (Artículo 14, Reglamento (UE) 2024/1689).

Por ejemplo, un algoritmo que evalúe el desempeño de los empleados debe permitir la intervención de un supervisor humano para revisar decisiones críticas, como despidos o promociones.

• Documentación y trazabilidad:

> *"Los proveedores de sistemas de IA de alto riesgo deberán mantener registros sobre el funcionamiento del sistema"* (Artículo 12, Reglamento (UE) 2024/1689).

Las empresas deben documentar cómo se entrenó el algoritmo, qué datos se utilizaron y cómo se toman las decisiones, facilitando auditorías regulatorias.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

• ***Ciberseguridad:***

> *"Los sistemas de IA de alto riesgo deberán ser resistentes a intentos de alterar su uso, comportamiento o rendimiento"* (Artículo 15, Reglamento (UE) 2024/1689).

Esto incluye proteger los sistemas contra ataques que puedan manipular resultados, como la alteración de datos en un sistema de selección automatizada.

Como ejemplo práctico, una empresa que utilice IA para asignar turnos laborales debe garantizar que el algoritmo sea transparente, utilice datos representativos y permita la supervisión humana para evitar decisiones injustas, como la asignación desproporcionada de turnos nocturnos a ciertos grupos.

Cabe destacar las siguientes sentencias, que han interpretado ciertos aspectos normativos:

o **Sentencia del TJUE,** C-311/18 (Schrems II, 16 de julio de 2020): Aunque centrada en transferencias internacionales de datos, esta sentencia subrayó la necesidad de proteger los derechos fundamentales en sistemas automatizados, un principio aplicable a la IA en Recursos Humanos.

o Sentencia del TJUE, C-203/15 (Tele2 Sverige, 21 de diciembre de 2016): Estableció que el tratamiento masivo de datos personales debe ser proporcional y justificado, un precedente para los sistemas de IA que procesan grandes volúmenes de datos de empleados.

c) Prohibiciones de sistemas de riesgo inaceptable

El AI Act prohíbe sistemas de inteligencia artificial que impliquen riesgos inaceptables, especialmente en el ámbito laboral:

> *"Se prohíben las prácticas de inteligencia artificial que impliquen un riesgo inaceptable, como los sistemas que utilicen técnicas subliminales, exploten vulnerabilidades de grupos específicos o realicen puntuaciones sociales"* (Artículo 5, Reglamento (UE) 2024/1689).

Su aplicación en RR. HH. se concreta en estos principios:

• **Reconocimiento de emociones:** Prohibido en el entorno laboral debido a su potencial para vulnerar la privacidad y los derechos fundamentales. Por ejemplo, un sistema que analice expresiones faciales para evaluar el estado emocional de un candidato durante una entrevista está prohibido.

• **Puntuación social:** Prohibido por su riesgo de discriminación y vigilancia masiva. Esto incluye sistemas que asignen puntuaciones a empleados basadas en métricas subjetivas, como el "nivel de compromiso".

Como ejemplo práctico, una empresa que intente implementar un sistema de IA para evaluar el "entusiasmo" de los empleados mediante análisis de voz o gestos sería sancionada, ya que estos sistemas infringen los derechos fundamentales.

d) Obligaciones de transparencia

Los sistemas de IA generativa, como chatbots para entrevistas o herramientas de creación de contenidos formativos, deben cumplir con requisitos específicos de transparencia.

> *"Los proveedores de sistemas de IA generativa deberán garantizar que los contenidos generados por dichos sistemas sean identificables como tales"* (Artículo 52, Reglamento (UE) 2024/1689).

Su aplicación en RR. HH. comprende la siguiente aplicación:

- Los empleados y candidatos deben saber cuándo interactúan con IA, por ejemplo, en entrevistas automatizadas o sistemas de formación personalizada.
- Los proveedores deben cumplir con normas de derechos de autor, asegurando que los datos de entrenamiento no infrinjan la propiedad intelectual (Samuelson, 2020).
- Los proveedores deben notificar incidentes graves, como fallos en el sistema que afecten decisiones laborales, en un plazo de 72 horas.

e) Oficina europea de IA

La Oficina europea de IA, establecida por el AI Act, coordina la implementación del reglamento en los Estados miembros.

> *"La Oficina de Inteligencia Artificial de la Unión Europea tendrá como objetivo coordinar la aplicación del presente Reglamento en los Estados miembros, proporcionando orientación técnica y supervisando el cumplimiento"* (Artículo 56, Reglamento (UE) 2024/1689).

Sus principales funciones son:

- Proporcionar directrices técnicas para sistemas de RR. HH.
- Coordinar con organismos nacionales, como la AESIA en España.
- Supervisar el cumplimiento de los sistemas de alto riesgo.

La Oficina Europea de IA puede emitir directrices específicas para sistemas de selección automatizada, recomendando auditorías periódicas y supervisión humana.

2.1.3 Reglamento de Servicios Digitales

El Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales).

a) Desafíos legales y éticos

La implementación de inteligencia artificial (IA) en la gestión de recursos humanos ha transformado procesos como la contratación, la evaluación del desempeño y la gestión del talento.

Su uso plantea desafíos legales y éticos, especialmente en el marco de la Unión Europea (UE) con la

Financiado por la Unión Europea NextGenerationEU · Gobierno de España · Ministerio de Derechos Sociales y Agenda 2030 · Plan de Recuperación, Transformación y Resiliencia · Comunidad de Madrid · Dirección General de Evaluación, Calidad e Innovación · Consejería de Familia, Juventud y Asuntos Sociales

entrada en vigor del Digital Services Act (DSA), pues la aplicación de la IA en RR. HH. requiere cumplir los principios de transparencia, responsabilidad y protección de datos.

El Digital Services Act (Reglamento (UE) 2022/2065), implementado en noviembre de 2022 y plenamente aplicable desde febrero de 2024, establece un marco normativo para los servicios digitales en la UE, incluyendo plataformas que utilizan IA.

Aunque el DSA se centra principalmente en intermediarios digitales, sus principios de transparencia, responsabilidad y protección de derechos fundamentales tienen implicaciones directas para el uso de la IA en RR. HH. El DSA, junto con otras normativas como el Reglamento General de Protección de Datos, configura el marco legal para la implementación de sistemas de inteligencia artificial en procesos de RR. HH.

El DSA tiene como objetivo regular los servicios digitales para garantizar un entorno seguro, transparente y responsable. Aunque no aborda específicamente la IA en el caso de la gestión de personas, sus disposiciones sobre algoritmos, moderación de contenidos y responsabilidad de las plataformas son relevantes para los sistemas de IA utilizados en la selección de personal, evaluación de empleados y otras funciones de RR. HH.

b) Principales disposiciones aplicables

Las principales disposiciones aplicables incluyen:

- **Transparencia en los sistemas algorítmicos**

El artículo 24 del DSA exige que las plataformas expliquen cómo funcionan sus sistemas de recomendación algorítmica.

Así se destaca en el Considerando 84:

> *Al evaluar dichos riesgos sistémicos, los prestadores de plataformas en línea de muy gran tamaño y de motores de búsqueda en línea de muy gran tamaño deben centrarse en los sistemas u otros elementos que puedan contribuir a los riesgos, incluidos todos los sistemas algorítmicos que puedan ser pertinentes, en particular sus sistemas de recomendación y de publicidad, dedicando atención a las prácticas conexas de recogida y uso de datos. También deben evaluar si sus condiciones generales y su cumplimiento son adecuados, así como sus procesos de moderación de contenidos, herramientas técnicas y recursos asignados. Al evaluar los riesgos sistémicos identificados en el presente Reglamento, dichos prestadores también deben centrarse en la información que no sea ilícita, pero que contribuya a los riesgos sistémicos identificados en el presente Reglamento. Por consiguiente, dichos prestadores deben dedicar especial atención a cómo se utilizan sus servicios para difundir o amplificar contenidos incorrectos o engañosos, incluida la desinformación. Cuando la amplificación algorítmica de la información contribuya a los riesgos sistémicos, dichos prestadores deben reflejarlo debidamente en sus evaluaciones de riesgos. (...)*

En Recursos Humanos esto implica que las herramientas de inteligencia artificial deben proporcionar información clara sobre cómo se toman decisiones, como la selección de candidatos o la evaluación del desempeño.

• Responsabilidad por decisiones automatizadas

El artículo 22 del RGPD, ya citado, complementado por el DSA, prohíbe decisiones exclusivamente automatizadas que tengan efectos legales o significativos, a menos que se cuente con el consentimiento explícito del usuario o exista una base legal.

• Protección de datos personales

El DSA refuerza las obligaciones del RGPD en cuanto a la privacidad y seguridad de los datos, especialmente en sistemas de IA que procesan información sensible de empleados o candidatos, para evitar la vigilancia excesiva y respeten los derechos de privacidad.

Los sistemas de IA utilizados en el filtrado de currículums o entrevistas automatizadas deben cumplir con los principios de no discriminación y transparencia. Por ejemplo, el uso de algoritmos para evaluar candidatos debe evitar sesgos que puedan violar la Directiva de Igualdad de Trato (2000/78/CE). Además, el DSA exige que los empleadores informen a los candidatos sobre el uso de herramientas de IA y los criterios de decisión.

Las herramientas de IA que supervisen el rendimiento de los empleados deben garantizar que los datos recopilados (como métricas de productividad) sean procesados de manera transparente y con el consentimiento del empleado.

En la gestión del talento, la IA puede identificar oportunidades de desarrollo o predecir la rotación de empleados. Sin embargo, el DSA requiere que las organizaciones sean responsables de las decisiones basadas en estas predicciones, asegurando que no se generen consecuencias negativas injustas para los trabajadores.

Estos son los principales desafíos:

- **Sesgos algorítmicos:** Los sistemas de IA pueden perpetuar sesgos si los datos de entrenamiento no son representativos o están sesgados.
- **Falta de transparencia:** Muchos sistemas de IA operan como cajas negras, lo que dificulta cumplir con las exigencias de transparencia del DSA.
- **Cumplimiento normativo:** Las organizaciones deben alinear sus sistemas de IA con múltiples regulaciones (DSA, RGPD, directivas de igualdad), lo que puede ser complejo.

Como recomendaciones aplicables, cabe mencionar:

- **Auditorías de IA:** Realizar auditorías periódicas para identificar y mitigar sesgos en los algoritmos de RR. HH.
- **Transparencia proactiva:** Informar a empleados y candidatos sobre el uso de IA, incluyendo los criterios y procesos de decisión.
- **Capacitación:** Formar al personal de Recursos Humanos en el uso ético de los sistemas de inteligencia artificial y las obligaciones legales del DSA y RGPD.
- **Consentimiento informado**: Obtener el consentimiento explícito de los empleados para el procesamiento de datos en sistemas de IA.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

El Digital Services Act proporciona un marco robusto para garantizar que la implementación de la IA en RRHH sea transparente, responsable y respetuosa con los derechos fundamentales. Las organizaciones deben adoptar un enfoque proactivo para cumplir con las disposiciones del DSA y otras normativas relevantes, como el RGPD. Al hacerlo, no solo mitigan riesgos legales, sino que también fomentan la confianza de empleados y candidatos en el uso de la IA. La interacción entre el DSA y el Reglamento de Inteligencia Artificial (AI Act) seguirá adaptando y moldeando las prácticas de RRHH.

c) Principios clave de la DSA en entornos laborales

La DSA establece un marco legal para un entorno digital más seguro, transparente y justo.

Tabla 15: *Principios de la Ley de Servicios Digitales que afectan a la gestión interna de las empresas y a la protección de los trabajadores en entornos laborales digitales*

Principio Clave	Descripción	Aplicación entornos laborales	Artículo DSA
Transparencia	Las plataformas deben informar claramente sobre sus algoritmos, decisiones de moderación y publicidad	Empleados y empleadores conocen cómo se gestionan los datos y decisiones automatizadas	Art. 14, 15, 24, 27, 42
Protección de derechos fundamentales	Garantiza la libertad de expresión, privacidad y no discriminación en línea	Salvaguarda de la privacidad y derechos de los trabajadores en plataformas digitales	Art. 3, 7, 8, 14
Responsabilidad de las plataformas	Obligación de eliminar contenido ilegal de forma rápida y efectiva tras notificación	Mecanismos claros para denunciar y retirar contenidos ilícitos relacionados con el trabajo	Art. 16, 17, 19, 20
Evaluación y gestión de riesgos	Auditorías y evaluaciones periódicas sobre riesgos para los derechos y la seguridad	Prevención de riesgos para empleados en el uso de herramientas digitales	Art. 34, 35, 36
Prohibición de patrones oscuros	Se prohíben prácticas engañosas que manipulan decisiones de los usuarios	Evita manipulaciones en plataformas internas y externas de la empresa	Art. 25
Protección especial a menores	Interfaz y publicidad adaptadas para proteger a menores; no se permite segmentación por perfil	Relevante en empresas con trabajadores menores de edad o en formación	Art. 28, 39
Acceso a vías de reclamación	Mecanismos accesibles para notificar y reclamar decisiones sobre contenido y datos	Empleados pueden reclamar decisiones automatizadas o sobre sus datos personales	Art. 20, 21, 53
Actualización y formación	Capacitación obligatoria para el personal sobre nuevas obligaciones y procedimientos	Formación continua en cumplimiento y uso responsable de plataformas digitales	Art. 45, 47
Monitoreo y auditoría	Sistemas de control interno y auditorías independientes para asegurar el cumplimiento	Seguimiento del cumplimiento de la DSA en procesos internos y externos	Art. 37, 40, 42

Fuente: elaboración propia.

2.1.4 Data Governance Act

La Data Governance Act (DGA), el Reglamento (UE) 2022/868 del Parlamento Europeo y del Consejo de 30 de mayo de 2022 relativo a la gobernanza de los datos, regula mecanismos para facilitar la reutilización segura de determinados datos protegidos, especialmente aquellos en poder de organismos públicos, y fomenta la creación de servicios de intermediación de datos bajo estrictos requisitos de transparencia y protección.

La DGA, en el ámbito de la IA y su aplicación a los RR. HH., permite el acceso y la reutilización de datos laborales y de recursos humanos bajo condiciones controladas, de modo que facilita el desarrollo de sistemas de IA más robustos y representativos.

Sin embargo, establece salvaguardias para la protección de datos personales y sensibles, asegurando que cualquier reutilización o intercambio de datos en procesos de selección, evaluación o gestión del talento se realice bajo estándares éticos y legales. Así, promueve la creación de entidades intermediarias certificadas que faciliten el intercambio de datos entre empresas, organismos públicos y terceros, para preservar la confianza y la trazabilidad en el uso de datos para IA en RR. HH.

Esta norma refuerza la transparencia y la supervisión sobre el uso de datos, lo que resulta esencial para prevenir sesgos y proteger los derechos fundamentales en la aplicación de IA en la gestión de personas.

Esta normativa, al igual que la posterior Data Act, que se analiza seguidamente, refuerza la transparencia, la equidad y la gobernanza responsable de los datos en el uso de IA en recursos humanos, alineándose con el marco europeo de protección de datos y ética digital

2.1.5 Data Act

El Reglamento (UE) 2023/2854 del Parlamento Europeo y del Consejo de 13 de diciembre de 2023 sobre normas armonizadas para el acceso justo y el uso de los datos (Data Act), establece un marco armonizado para el acceso, uso y compartición de datos generados por productos conectados y servicios digitales, tanto entre empresas como entre empresas y consumidores, en cuanto que no sean propiamente datos personales.

En el contexto de los RR. HH., esta regulación garantiza que los datos generados por dispositivos y plataformas en el entorno laboral (por ejemplo, sistemas de control de presencia, plataformas de evaluación de desempeño o herramientas de selección automatizada) sean accesibles tanto para los empleados como para la organización.

Adicionalmente, supone un refuerzo del derecho de portabilidad y acceso a los datos contemplado en el RGPD, pues permite a los empleados solicitar y transferir sus datos entre diferentes servicios o empleadores, lo que favorece la transparencia y la movilidad laboral.

Respecto de las empresas, impone la obligación de gestionar los datos bajo principios de equidad, transparencia y no discriminación, a fin de prevenir abusos contractuales y desequilibrios en el acceso a la información.

En el marco de la IA, esta norma facilita la compartición de datos con terceros (por ejemplo,

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

proveedores de inteligencia artificial o plataformas de intermediación laboral) bajo condiciones justas y seguras, siempre con respeto a los derechos de los titulares de los datos.

Además, establece obligaciones específicas para la protección de datos sensibles y para la interoperabilidad de sistemas, lo que es clave en la integración de soluciones de inteligencia artificial en Recursos Humanos.

2.1.6 Reglamento de Ejecución

Por el Reglamento de Ejecución (UE) 2023/138 de la Comisión de 21 de diciembre de 2022 se establecen una lista de conjuntos de datos específicos de alto valor y modalidades de publicación y reutilización. Se centra en la publicación y reutilización de conjuntos de datos de alto valor del sector público, pero sus principios de apertura, transparencia y estandarización de datos son relevantes para la gestión de datos en RR. HH.:

a) Disponibilidad y reutilización de datos internos

Este Reglamento no obliga directamente a las empresas privadas a publicar sus datos de RR. HH., pero sí promueve una cultura de datos abiertos y reutilizables. Esto puede influir en cómo las organizaciones gestionan internamente sus propios datos de los empleados. Para desarrollar e implementar IA ética en RR. HH., es fundamental que las empresas puedan acceder y reutilizar sus propios datos de manera eficiente y segura.

b) Calidad y estandarización de datos

El énfasis del Reglamento en la calidad y estandarización de los conjuntos de datos de alto valor subraya la importancia de contar con datos limpios, consistentes y bien estructurados para cualquier aplicación de IA. En RR. HH., esto se traduce en la necesidad de bases de datos de empleados, rendimiento, habilidades o procesos de selección que sean fiables y interoperables, minimizando sesgos introducidos por datos inconsistentes.

c) Gobernanza de datos reforzada

La exigencia de modalidades de publicación y reutilización para datos de alto valor implica una gobernanza de datos rigurosa. En RR. HH., esto se traduce en la necesidad de establecer políticas claras para la recolección, almacenamiento, acceso y uso de datos personales, asegurando el cumplimiento no solo del RGPD, sino también de principios de transparencia en el manejo de la información.

d) Fomento de la innovación responsable

Al promover la reutilización de datos, el Reglamento busca impulsar la innovación. En el contexto laboral y de Recursos Humanos, esto podría significar el desarrollo de soluciones de inteligencia artificial más sofisticadas y éticas al disponer de conjuntos de datos de alta calidad. Sin embargo, la reutilización en Recursos Humanos debe siempre ser ponderada con la privacidad y los derechos de los individuos.

2.1.7 Directiva (UE)

Ha de tenerse en cuenta 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. La Directiva (UE) 2024/2831, aprobada el 23 de octubre de 2024, tiene como objetivo mejorar las condiciones laborales y la protección de datos de los trabajadores de plataformas digitales en la Unión Europea.

Con un plazo de transposición hasta el 2 de diciembre de 2026, la directiva abarca a todos los trabajadores de plataformas, incluidos aquellos que operan a través de intermediarios, asegurando equidad y protección en la economía digital.

Establece una presunción legal de relación laboral cuando se detectan indicios de control y dirección por parte de la plataforma, promoviendo derechos laborales y negociación colectiva.

Un aspecto clave es su regulación de la inteligencia artificial, al exigir transparencia en los sistemas algorítmicos que gestionan tareas, horarios o evaluaciones, y requerir supervisión humana en decisiones automatizadas para evitar sesgos o discriminación. Además, prohíbe el uso de datos sensibles y garantiza que las plataformas informen sobre el funcionamiento de sus algoritmos de inteligencia artificial.

La gestión algorítmica se refiere al uso de sistemas automatizados que emplean datos de seguimiento y otra información para organizar, asignar, controlar, supervisar y evaluar el trabajo de las personas que prestan servicios en plataformas digitales. Esta directiva dedica un capítulo específico a esta materia y establece limitaciones claras y garantías para proteger a los trabajadores frente a los riesgos derivados del uso de IA en la gestión laboral.

Este marco normativo europeo sitúa la gestión algorítmica en el centro del debate sobre la protección laboral en la era digital, pues de partida impone límites claros al uso de IA en la gestión de personas para evitar discriminación, invasión de la privacidad y decisiones opacas. La exigencia de transparencia, supervisión humana y participación sindical contribuye a equilibrar la eficiencia tecnológica con la defensa de los derechos fundamentales de los trabajadores, especialmente en sectores vulnerables como el trabajo en plataformas.

2.1.8 Directiva sobre responsabilidad en IA

La Propuesta de Directiva sobre responsabilidad civil extracontractual por daños causados por IA (COM/2022/496) establece un marco para reclamaciones por daños causados por sistemas de inteligencia artificial.

> *"Los Estados miembros garantizarán que las personas físicas o jurídicas que sufran daños causados por un sistema de IA puedan obtener una compensación adecuada"* (Artículo 4, COM/2022/496).

Los candidatos o empleados perjudicados por decisiones algorítmicas discriminatorias, como la exclusión injusta en un proceso de selección, pueden reclamar compensaciones. La directiva facilita la carga de la prueba, exigiendo a los proveedores que proporcionen información sobre el funcionamiento del sistema.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

2.1.9 Pacto por la IA

El Pacto por la IA es una iniciativa voluntaria que invita a los proveedores e implementadores a cumplir con las disposiciones del AI Act antes de su entrada en vigor.

> *"El Pacto por la Inteligencia Artificial invita a los proveedores e implementadores a cumplir voluntariamente con las disposiciones del Reglamento de IA antes de su entrada en vigor"* (Comunicación de la Comisión Europea, 2024).

Este Pacto fomenta buenas prácticas en RR. HH., como la transparencia, la auditoría de sesgos y la supervisión humana. Por ejemplo, una empresa que participe en el Pacto puede comprometerse a auditar sus algoritmos de selección anualmente para garantizar la equidad.

2.1.10 Reglamento de ePrivacy

El Reglamento ePrivacy, aún en fase de propuesta y sin fecha definitiva de entrada en vigor, busca reforzar la privacidad en las comunicaciones electrónicas en la Unión Europea, derogando la Directiva 2002/58/CE.

Como lex specialis del RGPD, establece normas específicas para proteger datos y metadatos de comunicaciones electrónicas, exigiendo consentimiento expreso, libre e informado para su tratamiento, incluyendo cookies y tecnologías de seguimiento.

En relación con la inteligencia artificial, regula los sistemas algorítmicos utilizados en el procesamiento automatizado de datos, exigiendo transparencia sobre su funcionamiento y prohibiendo prácticas invasivas, como los "muros de cookies", que puedan derivar de decisiones basadas en IA. Abarca a proveedores de servicios como WhatsApp o Skype Empresarial y comunicaciones máquina a máquina (Internet de las Cosas), asegurando que las aplicaciones de IA respeten los derechos de privacidad. Con sanciones de hasta 20 millones de euros o el 4 % de la facturación anual, promueve un mercado digital seguro y equitativo, con un plazo de transposición de 24 meses tras su publicación.

2.2 Marco legal de España

En España, la regulación de la inteligencia artificial en Recursos Humanos combina normativas nacionales y europeas, centrándose en la protección de datos, la transparencia, la no discriminación y la supervisión humana.

Las principales normativas incluyen el Reglamento General de Protección de Datos (RGPD), la Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales (LOPDGDD), la Carta de Derechos Digitales o la Ley de Propiedad Intelectual.

2.2.1 Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales (LOPDGDD)

La LOPDGDD, publicada en el Boletín Oficial del Estado (BOE) el 6 de diciembre de 2018 y en vigor desde el 7 de diciembre de 2018, complementa el RGPD en el contexto español, al introducir disposiciones específicas que afectan el uso de tecnologías como la inteligencia artificial en Recursos Humanos.

Esta ley no solo adapta la normativa europea, sino que, sin tener un precedente en la Unión Europea, también incorpora un catálogo de derechos digitales que son especialmente relevantes en el ámbito laboral. A continuación, se detallan aquellos aspectos clave de la LOPDGDD relacionados con la implementación de IA en RR. HH.

a) Principios Fundamentales de la LOPDGDD en el contexto de la IA

La LOPDGDD establece principios que deben regir cualquier tratamiento de datos personales, incluyendo aquellos realizados mediante sistemas de IA. Estos principios, alineados con el RGPD, son fundamentales para garantizar un uso ético y legal de la IA en RR. HH.:

• Licitud, lealtad y transparencia (Art. 2 LOPDGDD):

> *1. Lo dispuesto en los Títulos I a IX y en los artículos 89 a 94 de la presente ley orgánica se aplica a cualquier tratamiento total o parcialmente automatizado de datos personales, así como al tratamiento no automatizado de datos personales contenidos o destinados a ser incluidos en un fichero.*
>
> *2. Esta ley orgánica no será de aplicación:*
>
> > *a) A los tratamientos excluidos del ámbito de aplicación del Reglamento general de protección de datos por su artículo 2.2, sin perjuicio de lo dispuesto en los apartados 3 y 4 de este artículo.*
> > *b) A los tratamientos de datos de personas fallecidas, sin perjuicio de lo establecido en el artículo 3.*
> > *c) A los tratamientos sometidos a la normativa sobre protección de materias clasificadas.*
>
> *3. Los tratamientos a los que no sea directamente aplicable el Reglamento (UE) 2016/679 por afectar a actividades no comprendidas en el ámbito de aplicación del Derecho de la Unión Europea, se regirán por lo dispuesto en su legislación específica si la hubiere y supletoriamente por lo establecido en el citado reglamento y en la presente ley orgánica. Se encuentran en esta situación, entre otros, los tratamientos realizados al amparo de la legislación orgánica del régimen electoral general, los tratamientos realizados en el ámbito de instituciones penitenciarias y los tratamientos derivados del Registro Civil, los Registros de la Propiedad y Mercantiles.*
>
> El artículo 64 de Ley Orgánica 3/2018, modificado por Real Decreto 817/2023, para el caso concreto de los trabajadores dispone el deber de información en estos términos:
>
> > *"Los trabajadores tendrán derecho a ser informados por el empresario sobre los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones laborales, el acceso o mantenimiento del empleo, incluida la elaboración de perfiles".*
>
> Las empresas deben informar a los empleados y candidatos sobre el uso de herramientas de IA, incluyendo las finalidades del tratamiento, la base jurídica (como el consentimiento

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

explícito o el interés legítimo) y los derechos asociados. Por ejemplo, si se utiliza IA para analizar currículums o evaluar entrevistas, los candidatos deben ser informados de manera clara y accesible.

• Limitación de la finalidad (Art. 6 LOPDGDD):

> *1. Cuando se pretenda fundar el tratamiento de los datos en el consentimiento del afectado para una pluralidad de finalidades será preciso que conste de manera específica e inequívoca que dicho consentimiento se otorga para todas ellas.*
>
> *2. No podrá supeditarse la ejecución del contrato a que el afectado consienta el tratamiento de los datos personales para finalidades que no guarden relación con el mantenimiento, desarrollo o control de la relación contractual.*

Los datos tratados mediante IA deben utilizarse únicamente para los fines especificados, como procesos de selección, evaluación de desempeño o gestión de talento.

• Minimización de datos:

Solo se deben recopilar los datos estrictamente necesarios para el propósito del tratamiento.

Por ejemplo, un sistema de IA para selección de personal no debe recopilar datos irrelevantes como opiniones políticas o información de salud, salvo que sea estrictamente necesario y esté justificado.

• Exactitud (Art. 4 LOPDGDD):

> *1. Conforme al artículo 5.1.d) del Reglamento (UE) 2016/679 los datos serán exactos y, si fuere necesario, actualizados.*

Los sistemas de IA deben garantizar que los datos tratados sean precisos y actualizados para evitar decisiones basadas en información errónea, lo que podría derivar en sesgos o discriminación.

• Limitación del plazo de conservación (Art. 24 LOPDGDD):

> *4. Los datos de quien formule la comunicación y de los empleados y terceros deberán conservarse en el sistema de denuncias únicamente durante el tiempo imprescindible para decidir sobre la procedencia de iniciar una investigación sobre los hechos denunciados.*
>
> *En todo caso, transcurridos tres meses desde la introducción de los datos, deberá procederse a su supresión del sistema de denuncias, salvo que la finalidad de la conservación sea dejar evidencia del funcionamiento del modelo de prevención de la comisión de delitos por la persona jurídica. Las denuncias a las que no se haya dado curso solamente podrán constar de forma anonimizada, sin que sea de aplicación la obligación de bloqueo prevista en el artículo 32 de esta ley orgánica.*

Los datos deben conservarse solo durante el tiempo necesario para cumplir con la finalidad del tratamiento, salvo obligaciones legales que exijan su conservación.

• Seguridad de los datos (Art. 10 y 22 LOPDGDD, entre otros):

El uso de la IA en el ámbito de los Recursos Humanos exige igualmente el cumplimiento de las necesarias medidas de seguridad.

Se pueden destacar a título ilustrativo dos ámbitos relevantes, uno más extraordinario, como los procesos penales; y otro más cotidiano, como la videovigilancia en el entorno laboral:

> *Artículo 10. Tratamiento de datos de naturaleza penal.*
>
> *1. El tratamiento de datos personales relativos a condenas e infracciones penales, así como a procedimientos y medidas cautelares y de seguridad conexas, para fines distintos de los de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, solo podrá llevarse a cabo cuando se encuentre amparado en una norma de Derecho de la Unión, en esta ley orgánica o en otras normas de rango legal (...).*
>
> *3. Fuera de los supuestos señalados en los apartados anteriores, los tratamientos de datos referidos a condenas e infracciones penales, así como a procedimientos y medidas cautelares y de seguridad conexas solo serán posibles cuando sean llevados a cabo por abogados y procuradores y tengan por objeto recoger la información facilitada por sus clientes para el ejercicio de sus funciones.*
>
> *Artículo 22. Tratamientos con fines de videovigilancia.*
>
> *1. Las personas físicas o jurídicas, públicas o privadas, podrán llevar a cabo el tratamiento de imágenes a través de sistemas de cámaras o videocámaras con la finalidad de preservar la seguridad de las personas y bienes, así como de sus instalaciones.*
>
> *2. Solo podrán captarse imágenes de la vía pública en la medida en que resulte imprescindible para la finalidad mencionada en el apartado anterior (...).*
>
> *3. Los datos serán suprimidos en el plazo máximo de un mes desde su captación, salvo cuando hubieran de ser conservados para acreditar la comisión de actos que atenten contra la integridad de personas, bienes o instalaciones. En tal caso, las imágenes deberán ser puestas a disposición de la autoridad competente en un plazo máximo de setenta y dos horas desde que se tuviera conocimiento de la existencia de la grabación (...).*

Las empresas deben implementar medidas técnicas y organizativas para garantizar la confidencialidad, integridad y disponibilidad de los datos tratados por sistemas de inteligencia artificial, conforme al Art. 32 del RGPD y reforzado por la LOPDGDD.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

b) Derechos digitales en el ámbito laboral

El Título X de la LOPDGDD introduce un conjunto de derechos digitales que son especialmente relevantes para el uso de IA en Recursos Humanos, ya que protegen la privacidad y los derechos de los trabajadores en un entorno digital. Los más relevantes son:

- **Derecho a la intimidad y uso de dispositivos digitales en el ámbito laboral (Art. 87 LOPDGDD):**

> *1. Los trabajadores y los empleados públicos tendrán derecho a la protección de su intimidad en el uso de los dispositivos digitales puestos a su disposición por su empleador.*
>
> *2. El empleador podrá acceder a los contenidos derivados del uso de medios digitales facilitados a los trabajadores a los solos efectos de controlar el cumplimiento de las obligaciones laborales o estatutarias y de garantizar la integridad de dichos dispositivos.*
>
> *3. Los empleadores deberán establecer criterios de utilización de los dispositivos digitales respetando en todo caso los estándares mínimos de protección de su intimidad de acuerdo con los usos sociales y los derechos reconocidos constitucional y legalmente. En su elaboración deberán participar los representantes de los trabajadores.*
>
> *El acceso por el empleador al contenido de dispositivos digitales respecto de los que haya admitido su uso con fines privados requerirá que se especifiquen de modo preciso los usos autorizados y se establezcan garantías para preservar la intimidad de los trabajadores, tales como, en su caso, la determinación de los períodos en que los dispositivos podrán utilizarse para fines privados.*
>
> *Los trabajadores deberán ser informados de los criterios de utilización a los que se refiere este apartado.*

Los empleados tienen derecho a que el uso de dispositivos digitales proporcionados por la empresa (como ordenadores o móviles) respete su intimidad. Las empresas deben establecer políticas claras de uso aceptable, especialmente cuando se emplean sistemas de IA para monitorear el uso de estos dispositivos (por ejemplo, análisis de productividad o seguimiento de correos electrónicos).

- **Derecho a la desconexión digital (Art. 88 LOPDGDD):**

> *1. Los trabajadores y los empleados públicos tendrán derecho a la desconexión digital a fin de garantizar, fuera del tiempo de trabajo legal o convencionalmente establecido, el respeto de su tiempo de descanso, permisos y vacaciones, así como de su intimidad personal y familiar.*
>
> *2. Las modalidades de ejercicio de este derecho atenderán a la naturaleza y objeto de la relación laboral, potenciarán el derecho a la conciliación de la actividad laboral y la vida personal y familiar y se sujetarán a lo establecido en la negociación*

colectiva o, en su defecto, a lo acordado entre la empresa y los representantes de los trabajadores (...).

Los trabajadores tienen derecho a no ser contactados fuera de su horario laboral, lo que incluye evitar el uso de sistemas de inteligencia artificial que generen notificaciones o monitorización fuera de dicho horario.

Este derecho busca proteger la salud mental y la conciliación laboral, especialmente en contextos de teletrabajo donde la inteligencia artificial puede ser utilizada para supervisión remota.

• Derecho a la intimidad frente al uso de videovigilancia y geolocalización (Art. 89 LOPDGDD):

1. Los empleadores podrán tratar las imágenes obtenidas a través de sistemas de cámaras o videocámaras para el ejercicio de las funciones de control de los trabajadores o los empleados públicos previstas, respectivamente, en el artículo 20.3 del Estatuto de los Trabajadores y en la legislación de función pública, siempre que estas funciones se ejerzan dentro de su marco legal y con los límites inherentes al mismo. Los empleadores habrán de informar con carácter previo, y de forma expresa, clara y concisa, a los trabajadores o los empleados públicos y, en su caso, a sus representantes, acerca de esta medida.

En el supuesto de que se haya captado la comisión flagrante de un acto ilícito por los trabajadores o los empleados públicos se entenderá cumplido el deber de informar cuando existiese al menos el dispositivo al que se refiere el artículo 22.4 de esta ley orgánica.

2. En ningún caso se admitirá la instalación de sistemas de grabación de sonidos ni de videovigilancia en lugares destinados al descanso o esparcimiento de los trabajadores o los empleados públicos, tales como vestuarios, aseos, comedores y análogos.

3. La utilización de sistemas similares a los referidos en los apartados anteriores para la grabación de sonidos en el lugar de trabajo se admitirá únicamente cuando resulten relevantes los riesgos para la seguridad de las instalaciones, bienes y personas derivados de la actividad que se desarrolle en el centro de trabajo y siempre respetando el principio de proporcionalidad, el de intervención mínima y las garantías previstas en los apartados anteriores. La supresión de los sonidos conservados por estos sistemas de grabación se realizará atendiendo a lo dispuesto en el apartado 3 del artículo 22 de esta ley.

La LOPDGDD regula el uso de sistemas de videovigilancia y geolocalización en el lugar de trabajo, que a menudo incorporan tecnologías de IA (como reconocimiento facial o seguimiento de ubicación). Las empresas deben informar a los trabajadores sobre la instalación y propósito de estos sistemas, garantizar su proporcionalidad y limitar su uso a fines específicos, como la seguridad laboral. Por ejemplo, el uso de IA para analizar imágenes de cámaras de seguridad debe cumplir con estrictas medidas de privacidad.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

La Sentencia 3192/2022 (26/07/2022) del Tribunal Supremo aborda la colisión entre el derecho a la intimidad (art. 18 CE) y el uso de videovigilancia en el ámbito laboral (art. 89 LOPDGDD). En un caso de despido disciplinario en el servicio doméstico, se validan grabaciones sin información previa como prueba, al cumplir los principios de idoneidad, necesidad y proporcionalidad, considerando la vulnerabilidad de la empleadora (discapacidad severa) y la protección de su hogar. La sentencia confirma la procedencia del despido, alineándose con la jurisprudencia del TEDH (López Ribalda II) y la STC 39/2016.

• Derecho a la no discriminación en decisiones automatizadas:

Aunque no se menciona explícitamente en la LOPDGDD, el Art. 22 del RGPD, aplicable en España, prohíbe decisiones basadas únicamente en tratamientos automatizados, como los realizados por sistemas de IA, si estas tienen efectos jurídicos o afectan significativamente al interesado, salvo que se cuente con consentimiento explícito, autorización legal o sea necesario para un contrato.

En RR. HH., esto implica que decisiones como contrataciones, despidos o promociones basadas en IA deben incluir supervisión humana para evitar sesgos algorítmicos.

• Disposición adicional decimoséptima - Tratamiento de datos en el ámbito laboral:

La Disposición adicional decimoséptima de la LOPDGDD amplia en su contenido: cuando se aplica en el contexto de la inteligencia artificial, regula el tratamiento de datos de salud en España, estableciendo un marco que armoniza con el Reglamento General de Protección de Datos.

Esta disposición permite el procesamiento de dichos datos para fines de investigación científica, estadística o de interés público, siempre que se garantice la confidencialidad, seguridad y minimización de datos. Aunque el consentimiento explícito del interesado es generalmente requerido, se contemplan excepciones para proteger intereses vitales, cumplir obligaciones legales o realizar estudios en beneficio de la salud pública.

Las administraciones públicas, centros sanitarios y entidades de investigación pueden tratar estos datos bajo estrictas medidas de seguridad, como la seudonimización, asegurando la protección de los derechos fundamentales de los interesados.

c) Obligaciones de las empresas en el uso de IA

La LOPDGDD impone una serie de obligaciones a las empresas que implementan IA en procesos de RR. HH. para garantizar el cumplimiento normativo:

• Designación de un Delegado de Protección de Datos (DPD) (Art. 34 LOPDGDD):

1. Los responsables y encargados del tratamiento deberán designar un delegado de protección de datos en los supuestos previstos en el artículo 37.1 del Reglamento (UE) 2016/679 y, en todo caso, cuando se trate de las siguientes entidades: (...) [se enumeran las entidades].

> *2. Los responsables o encargados del tratamiento no incluidos en el párrafo anterior podrán designar de manera voluntaria un delegado de protección de datos, que quedará sometido al régimen establecido en el Reglamento (UE) 2016/679 y en la presente ley orgánica.*

Las empresas que realicen tratamientos de datos a gran escala o de categorías especiales (como datos biométricos en entrevistas analizadas por IA) deben designar un DPD.

Este profesional supervisa el cumplimiento del RGPD y la LOPDGDD, asegurando que los sistemas de IA respeten los derechos de los trabajadores.

• Evaluación de Impacto en la Protección de Datos (EIPD) (Art. 35 RGPD):

> *1. Cuando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento realizará, antes del tratamiento, una evaluación del impacto de las operaciones de tratamiento en la protección de datos personales. Una única evaluación podrá abordar una serie de operaciones de tratamiento similares que entrañen altos riesgos similares.*
>
> *2. El responsable del tratamiento recabará el asesoramiento del delegado de protección de datos, si ha sido nombrado, al realizar la evaluación de impacto relativa a la protección de datos (...).*

Cuando el uso de IA implique un riesgo alto para los derechos y libertades de los interesados (por ejemplo, en perfilado automatizado para selección de personal), las empresas deben realizar una EIPD para identificar y mitigar riesgos, como posibles sesgos en los algoritmos.

• Transparencia y consentimiento (Art. 13-14 RGPD y Art. 11 LOPDGDD):

> *1. Cuando los datos personales sean obtenidos del afectado el responsable del tratamiento podrá dar cumplimiento al deber de información establecido en el artículo 13 del Reglamento (UE) 2016/679 facilitando al afectado la información básica a la que se refiere el apartado siguiente e indicándole una dirección electrónica u otro medio que permita acceder de forma sencilla e inmediata a la restante información (...).*
>
> *3. Cuando los datos personales no hubieran sido obtenidos del afectado, el responsable podrá dar cumplimiento al deber de información establecido en el artículo 14 del Reglamento (UE) 2016/679 facilitando a aquel la información básica señalada en el apartado anterior, indicándole una dirección electrónica u otro medio que permita acceder de forma sencilla e inmediata a la restante información. En estos supuestos, la información básica incluirá también: a) Las categorías de datos objeto de tratamiento. b) Las fuentes de las que procedieran los datos.*

Las empresas deben informar a empleados y candidatos sobre el uso de IA, incluyendo finalidades, datos tratados, destinatarios y derechos de privacidad (acceso, rectificación...)

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

En casos donde el tratamiento se base en el consentimiento, este debe ser explícito, informado y libre, especialmente en procesos como entrevistas automatizadas o análisis de perfiles.

• Notificación de brechas de seguridad (Art. 33 RGPD y Art. 15 LOPDGDD):

> *1. El derecho de supresión se ejercerá de acuerdo con lo establecido en el artículo 17 del Reglamento (UE) 2016/679.*
>
> *2. Cuando la supresión derive del ejercicio del derecho de oposición con arreglo al artículo 21.2 del Reglamento (UE) 2016/679, el responsable podrá conservar los datos identificativos del afectado necesarios con el fin de impedir tratamientos futuros para fines de mercadotecnia directa.*

En caso de una brecha de seguridad que comprometa datos tratados por sistemas de IA (por ejemplo, una filtración de currículums analizados por un algoritmo), las empresas deben notificar a la Agencia Española de Protección de Datos (AEPD) en un plazo de 72 horas y, si es probable que afecte a los derechos de los interesados, informar a los afectados.

• Registro de Actividades de Tratamiento (RAT) (Art. 30 RGPD):

> *Cada responsable y, en su caso, su representante llevarán un registro de las actividades de tratamiento efectuadas bajo su responsabilidad. Dicho registro deberá contener toda la información indicada a continuación: a) el nombre y los datos de contacto del responsable y, en su caso, del corresponsable, del representante del responsable, y del delegado de protección de datos; b) los fines del tratamiento; c) una descripción de las categorías de interesados y de las categorías de datos personales; d) las categorías de destinatarios a quienes se comunicaron o comunicarán los datos personales, incluidos los destinatarios en terceros países u organizaciones internacionales; e) en su caso, las transferencias de datos personales a un tercer país o una organización internacional, incluida la identificación de dicho tercer país u organización internacional y, en el caso de las transferencias indicadas en el artículo 49, apartado 1, párrafo segundo, la documentación de garantías adecuadas; f) cuando sea posible, los plazos previstos para la supresión de las diferentes categorías de datos; g) cuando sea posible, una descripción general de las medidas técnicas y organizativas de seguridad a que se refiere el artículo 32, apartado 1.*

Las empresas deben mantener un registro interno de las actividades de tratamiento de datos, incluyendo aquellas realizadas por sistemas de IA, detallando la finalidad, los datos tratados, los plazos de conservación y las medidas de seguridad.

d) Medidas de seguridad y prevención de sesgos

La LOPDGDD, en línea con el Art. 32 del RGPD, exige que las empresas implementen medidas técnicas y organizativas para proteger los datos tratados por sistemas de inteligencia artificial.

Esto incluye:

• **Cifrado y anonimización:** Los datos utilizados por sistemas de IA deben estar protegidos mediante técnicas como el cifrado o la anonimización para minimizar riesgos en caso de brechas de seguridad.

• **Auditorías periódicas:** Las empresas deben realizar auditorías regulares de los sistemas de IA para garantizar que no introduzcan sesgos o discriminación. Por ejemplo, un algoritmo de selección de personal debe ser evaluado para evitar prejuicios basados en género, edad u origen étnico.

• **Formación continua:** Los sistemas de IA deben ser entrenados con datos representativos y actualizados para evitar resultados discriminatorios. La LOPDGDD refuerza la necesidad de garantizar la equidad en los procesos automatizados.

e) Sanciones y consecuencias por incumplimiento

El incumplimiento de la LOPDGDD y el RGPD en el uso de IA en RR. HH. puede resultar en sanciones severas, reguladas en los Artículos 71-74 de la LOPDGDD y el Art. 83 del RGPD. Las infracciones se clasifican en:

• **Muy graves**: Hasta 20 millones de euros o el 4 % del volumen de negocio anual global, por ejemplo, tratamiento de datos sin base jurídica o falta de transparencia.
• **Graves**: Hasta 10 millones de euros o el 2 % del volumen de negocio anual global, como no realizar una EIPD cuando sea necesaria.
• **Leves**: Multas menores, pero aún significativas, por incumplimientos formales como no informar adecuadamente a los interesados.

Además, la AEPD puede imponer medidas correctivas, como la prohibición de usar ciertos sistemas de IA o la obligación de implementar cambios en los procesos.

f) Implicaciones específicas para la IA en RR. HH.

El uso de inteligencia artificial en el ámbito de los Recursos Humanos, como en procesos de selección, evaluación de desempeño o monitoreo de empleados, plantea desafíos específicos que la LOPDGDD aborda:

• **Procesos de selección automatizados:** Los sistemas de IA que analizan currículums o entrevistas deben garantizar la no discriminación y la transparencia. Los candidatos tienen derecho a una explicación sobre las decisiones automatizadas que les afecten (Art. 22 RGPD).

• **Supervisión de empleados:** El uso de IA para supervisar la productividad o el comportamiento (por ejemplo, mediante análisis de correos o geolocalización) debe respetar los derechos a la intimidad y la desconexión digital (Art. 87-89 LOPDGDD).

• **Datos biométricos:** Si la inteligencia artificial utiliza datos biométricos (como reconocimiento facial en entrevistas), se considera un tratamiento de categorías especiales de datos, lo que requiere medidas adicionales de seguridad y, en muchos casos, el consentimiento explícito (Art. 9 RGPD).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

• **Canales internos de denuncia:** La LOPDGDD regula los sistemas de denuncias internas (Art. 24), que pueden incorporar IA para analizar patrones de comportamiento. Estos sistemas deben garantizar la confidencialidad de los denunciantes y los datos tratados.

> *1. Será lícita la creación y mantenimiento de sistemas de información a través de los cuales pueda ponerse en conocimiento de una entidad de Derecho privado, incluso anónimamente, la comisión en el seno de la misma o en la actuación de terceros que contratasen con ella, de actos o conductas que pudieran resultar contrarios a la normativa general o sectorial que le fuera aplicable. Los empleados y terceros deberán ser informados acerca de la existencia de estos sistemas de información (...).*

g) Recomendaciones para el cumplimiento

Para garantizar el cumplimiento del marco legal en la implementación de IA en RR. HH., las empresas deben:

• **Realizar una EIPD** antes de implementar sistemas de IA que impliquen tratamientos de datos sensibles o a gran escala.

• **Designar un DPD** en los casos requeridos, asegurando que supervise los procesos de IA.

• **Informar a empleados y candidatos** sobre el uso de inteligencia artificial, proporcionando información clara y accesible sobre las finalidades, los datos tratados y los derechos disponibles.

• **Implementar medidas técnicas** como cifrado, anonimización y auditorías algorítmicas para prevenir sesgos y proteger los datos.

• **Establecer políticas internas** que regulen el uso de dispositivos digitales, la videovigilancia y la desconexión digital, alineadas con los Art. 87-89 de la LOPDGDD.

• **Formar al personal de Recursos Humanos en las implicaciones legales y éticas** del uso de la inteligencia artificial, incluyendo la prevención de sesgos y la protección de datos personales.

• **Mantener un RAT** actualizado que incluya todos los tratamientos de datos realizados mediante IA.

• **Garantizar la supervisión humana** en decisiones automatizadas significativas, como contrataciones o despidos, para cumplir con el Art. 22 del RGPD.

h) Principios clave del marco de derecho y ética digital

La Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales que se ha analizado en este subapartado contempla también los principios esenciales que deben guiar la implementación de IA en Recursos Humanos, garantizando tanto la legalidad como la ética digital en el entorno laboral actual.

Tabla 16: *Principios clave del marco de derechos y ética digital*

Principio Clave	Descripción
Transparencia	Las empresas deben informar claramente sobre el uso de IA en procesos de RR. HH., explicar el funcionamiento de los algoritmos y garantizar que las decisiones automatizadas sean comprensibles para los empleados
Consentimiento y Control de datos	Es obligatorio obtener el consentimiento explícito de los empleados para procesar sus datos personales y permitirles decidir sobre el uso de sus datos
No Discriminación y Equidad	Se prohíbe el uso de IA para tomar decisiones injustificadas o discriminatorias (por ejemplo, en despidos o promociones), y se exige que los algoritmos sean auditados para evitar sesgos
Supervisión humana	Toda decisión relevante tomada por IA debe estar supervisada por personas capacitadas; no se permite que la IA decida de manera autónoma en cuestiones críticas de RR. HH.
Privacidad y Seguridad de la información	Es imprescindible proteger la confidencialidad de los datos personales de empleados y candidatos, aplicando medidas de seguridad y limitando el acceso a la información
Derecho a la explicación	Los empleados tienen derecho a conocer y solicitar explicaciones sobre cómo se han tomado decisiones automatizadas que les afectan
Limitación de uso de datos biométricos y emocionales	Se restringe el uso de IA para reconocimiento facial, biométrico o emocional, salvo excepciones legales muy justificadas
Auditoría y Evaluación continua	Es necesario auditar de forma regular los sistemas de IA para garantizar su funcionamiento ético, detectar errores y corregir desviaciones
Capacitación y Cultura Ética	Las empresas deben formar a sus equipos en derechos digitales, ética digital y uso responsable de la IA, promoviendo una cultura de responsabilidad
Impacto positivo y Proporcionalidad	El uso de IA debe generar beneficios para la organización y los empleados, evitando impactos negativos desproporcionados en los derechos fundamentales.

Fuente: elaboración propia.

2.2.2 Estatuto de los Trabajadores

El Artículo 20 bis del Estatuto de los Trabajadores (ET) español (Real Decreto Legislativo 2/2015, de 23 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto de los Trabajadores), introducido por el Real Decreto-ley 32/2021, de 28 de diciembre, se centra en los derechos de los trabajadores en relación con el entorno digital, específicamente en la protección de su intimidad frente al uso de dispositivos digitales y el derecho a la desconexión digital.

Sin embargo, en el contexto de la implementación de inteligencia artificial en Recursos Humanos, este artículo no aborda directamente la regulación de la IA, pero sí establece un marco general para los derechos digitales que puede servir como base para su aplicación.

El Artículo 20 bis ET debe tenerse en cuenta en el marco legal para la implementación de inteligencia artificial en lo relativo a la gestión de personas, integrando normativas relevantes como el RIA, la Ley Orgánica 3/2018 de Protección de Datos Personales y Garantía de los Derechos Digitales

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

(LOPDGDD), y otras disposiciones nacionales y europeas, con un enfoque en las implicaciones prácticas y legales:

> *Los trabajadores tienen derecho a la intimidad en el uso de los dispositivos digitales puestos a su disposición por el empleador, a la desconexión digital y a la intimidad frente al uso de dispositivos de videovigilancia y geolocalización en los términos establecidos en la legislación vigente en materia de protección de datos personales y garantía de los derechos digitales.*

Este artículo se enmarca en el contexto de la transformación digital del trabajo y busca proteger a los trabajadores frente al uso indebido de herramientas tecnológicas y la presión de estar permanentemente disponibles.

a) Derecho a la intimidad en el entorno digital

Los trabajadores tienen derecho a la protección de su intimidad en el uso de dispositivos digitales puestos a su disposición por la empresa. El empleador solo puede acceder a los contenidos de los dispositivos digitales proporcionados (como ordenadores, móviles, etc.) para controlar el cumplimiento de las obligaciones laborales, siempre que se respeten los principios de proporcionalidad, necesidad y mínima intervención.

La empresa debe establecer previamente criterios claros sobre el uso de estos dispositivos, respetando los derechos fundamentales del trabajador, como la intimidad y la protección de datos personales.

Se prohíbe el uso de sistemas de vigilancia (como cámaras o grabaciones) para fines distintos a los establecidos, y siempre debe respetarse la normativa de protección de datos (RGPD).

b) Derecho a la desconexión digital

Los trabajadores tienen derecho a no estar conectados fuera de su horario laboral, salvo en casos excepcionales previstos por la ley o acuerdos específicos. Este derecho busca garantizar el respeto al tiempo de descanso, permisos y conciliación de la vida personal y laboral.

Las empresas deben elaborar una política interna de desconexión digital, en consulta con los representantes de los trabajadores, que defina las modalidades de ejercicio de este derecho y fomente medidas para evitar la conexión fuera del horario laboral.

c) Obligaciones empresariales

Las empresas deben garantizar que las políticas de uso de dispositivos y desconexión sean claras, transparentes y consensuadas con los representantes de los trabajadores. Deben evitar prácticas que vulneren la intimidad, como el monitoreo excesivo o injustificado de correos electrónicos, mensajes o actividad en dispositivos corporativos.

En casos de fuerza mayor o necesidades operativas justificadas, se pueden establecer excepciones al derecho de desconexión, pero siempre respetando los límites legales y los acuerdos colectivos.

d) Transparencia y supervisión humana

El Artículo 20 bis, junto con el Artículo 64.4.d del ET y el Reglamento de IA, exige información clara, ya que los trabajadores y sus representantes deben conocer los parámetros y reglas de los sistemas de IA que afectan sus condiciones laborales, así como intervención humana, pues las decisiones automatizadas (por ejemplo, despidos o promociones) deben contar con supervisión humana para evitar vulneraciones de derechos.

> *Artículo 64.4. El comité de empresa, con la periodicidad que proceda en cada caso, tendrá derecho a: (...) d) Ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles.*

e) Prevención de discriminación

Los sistemas de IA en RR. HH. pueden perpetuar sesgos históricos (por ejemplo, en la selección de candidatos). El Artículo 20 bis, al proteger la intimidad, se alinea con el principio de no discriminación (Artículo 17 del ET) y exige que los algoritmos sean auditados para garantizar equidad:

> *1. Se entenderán nulos y sin efecto los preceptos reglamentarios, las cláusulas de los convenios colectivos, los pactos individuales y las decisiones unilaterales del empresario que den lugar en el empleo, así como en materia de retribuciones, jornada y demás condiciones de trabajo, a situaciones de discriminación directa o indirecta desfavorables por razón de edad o discapacidad o a situaciones de discriminación directa o indirecta por razón de sexo, origen, incluido el racial o étnico, estado civil, condición social, religión o convicciones (...).*
>
> *Serán igualmente nulas las órdenes de discriminar y las decisiones del empresario que supongan un trato desfavorable de los trabajadores como reacción ante una reclamación efectuada en la empresa o ante una acción administrativa o judicial destinada a exigir el cumplimiento del principio de igualdad de trato y no discriminación.*

2.2.3 Modificación de la Ley del Estatuto de los Trabajadores

La Ley 12/2021, de 28 de septiembre, es por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales.

Esta norma responde a la necesidad de adaptar la legislación laboral al contexto actual, donde las plataformas digitales ejercen un control significativo sobre las condiciones de trabajo mediante gestión algorítmica, estableciendo una presunción de laboralidad para los repartidores.

La ley introduce una nueva disposición adicional en el Estatuto de los Trabajadores que presume la condición de laboral en las actividades de reparto o distribución de productos o mercancías cuando la empresa ejerce facultades de organización, dirección y control a través de plataformas digitales.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Esta presunción se fundamenta en la sentencia del Tribunal Supremo (STS 805/2020, de 25 de septiembre), que unificó doctrina y adaptó los criterios de dependencia y ajenidad al entorno digital, destacando que la aparente flexibilidad de los trabajadores puede ser ilusoria si está sujeta a consecuencias en su empleo o condiciones laborales. La norma no afecta lo dispuesto en el artículo 1.3 del Estatuto, manteniendo excepciones específicas:

> *1. Esta ley será de aplicación a los trabajadores que voluntariamente presten sus servicios retribuidos por cuenta ajena y dentro del ámbito de organización y dirección de otra persona, física o jurídica, denominada empleador o empresario.*
>
> *2. A los efectos de esta ley, serán empresarios todas las personas, físicas o jurídicas, o comunidades de bienes que reciban la prestación de servicios de las personas referidas en el apartado anterior, así como de las personas contratadas para ser cedidas a empresas usuarias por empresas de trabajo temporal legalmente constituidas.*
>
> *3. Se excluyen del ámbito regulado por esta ley: a) La relación de servicio de los funcionarios públicos (...). b) Las prestaciones personales obligatorias. c) La actividad que se limite, pura y simplemente, al mero desempeño del cargo de consejero o miembro de los órganos de administración (...). d) Los trabajos realizados a título de amistad, benevolencia o buena vecindad. e) Los trabajos familiares, salvo que se demuestre la condición de asalariados de quienes los llevan a cabo. Se considerarán familiares, a estos efectos, siempre que convivan con el empresario, el cónyuge, los descendientes, ascendientes y demás parientes por consanguinidad o afinidad, hasta el segundo grado inclusive y, en su caso, por adopción. f) La actividad de las personas que intervengan en operaciones mercantiles por cuenta de uno o más empresarios, siempre que queden personalmente obligados a responder del buen fin de la operación asumiendo el riesgo y ventura de la misma. g) En general, todo trabajo que se efectúe en desarrollo de relación distinta de la que define el apartado 1.*

El texto se apoya en la evolución del concepto de dependencia-subordinación, que ha pasado de una subordinación absoluta a una integración en el círculo organizativo y disciplinario de la empresa, según sentencias previas del Tribunal Supremo (como la STS de 11 de mayo de 1979).

Se reconoce que las plataformas digitales, al gestionar algorítmicamente el servicio, asumen riesgos y beneficios, ejerciendo un control indirecto que configura una relación laboral. Esto refuerza la aplicación del principio de realidad, priorizando la naturaleza real del vínculo laboral sobre la forma contractual, conforme al artículo 8.1 del Estatuto:

> *1. El contrato de trabajo se podrá celebrar por escrito o de palabra. Se presumirá existente entre todo el que presta un servicio por cuenta y dentro del ámbito de organización y dirección de otro y el que lo recibe a cambio de una retribución.*

Además, la ley busca garantizar la protección laboral en un contexto de transformación tecnológica, donde la gestión algorítmica redefine las relaciones laborales. La Inspección de Trabajo y Seguridad Social ha jugado un papel clave en la identificación de estas dinámicas, respaldando sentencias que han debatido la laboralidad de los repartidores. La norma enfatiza la necesidad de adaptar el derecho laboral a las nuevas formas de organización del trabajo, asegurando derechos y estabilidad a los trabajadores de plataformas digitales.

Esta legislación sustituye implícitamente al Real Decreto-ley 9/2021, de 11 de mayo, que abordaba el mismo objetivo, consolidando un marco normativo más estable y coherente. La Ley 12/2021 representa un paso hacia la modernización del derecho laboral español, alineándose con los principios de seguridad jurídica y protección de los trabajadores en la economía digital.

2.2.4 Agencia Española de Supervisión de Inteligencia Artificial (AESIA)

Creada por el Real Decreto 729/2023, de 19 de septiembre, la AESIA es el organismo encargado de supervisar el desarrollo y uso de la IA en España, con sede en A Coruña.

> *"La Agencia Española de Supervisión de Inteligencia Artificial tendrá como finalidad garantizar el cumplimiento de las disposiciones del Reglamento de la Unión Europea sobre inteligencia artificial, así como de las normativas nacionales aplicables"* (Artículo 3, Real Decreto 729/2023).

a) Funciones en RR. HH.

- Auditoría de sistemas de IA de alto riesgo, como los utilizados en selección, evaluación del desempeño o prevención de riesgos laborales.
- Supervisión del cumplimiento del RGPD, la LOPDGDD y el AI Act.
- Asesoramiento a empresas sobre el desarrollo ético de IA.

b) Impacto

Las empresas deben registrar sus sistemas de IA de alto riesgo en la AESIA y someterlos a evaluaciones periódicas. Por ejemplo, un sistema de IA que monitoree la productividad de los empleados debe ser auditado para garantizar que no vulnere la privacidad o genere discriminación.

Una empresa que utilice IA para predecir riesgos laborales debe demostrar a la AESIA que el sistema cumple con los principios de transparencia y no discrimina a trabajadores con discapacidades.

2.2.5 Carta de Derechos Digitales

La Carta de Derechos Digitales (2021), promovida por el Gobierno de España, establece un marco ético para el uso de la IA, inspirado en principios humanistas.

> *"Los ciudadanos tienen derecho a que los sistemas de inteligencia artificial utilizados en el ámbito laboral respeten los principios de igualdad, no discriminación y transparencia, garantizando la supervisión humana en las decisiones que afecten a sus derechos"* (Capítulo IV, Carta de Derechos Digitales).

Su aplicación en RR. HH. se concreta en estos aspectos:

- **Igualdad y no discriminación**: Prohíbe el uso de algoritmos que perpetúen sesgos.
- **Transparencia:** Exige que los empleados sean informados sobre el uso de IA en decisiones laborales.
- **Supervisión humana:** Garantiza que ninguna de las decisiones críticas sean completamente automatizadas, protegiendo la autonomía de los trabajadores.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030 Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

2.2.6 Ley de Propiedad Intelectual

La Ley de Propiedad Intelectual (Real Decreto Legislativo 1/1996, de 12 de abril, LPI) regula el uso de datos protegidos en el entrenamiento de modelos de IA, relevante para sistemas de IA generativa en RR. HH., como la creación de materiales formativos o evaluaciones personalizadas:

> *"Los derechos de autor protegen las obras literarias, artísticas o científicas expresadas en cualquier medio o soporte, tangible o intangible, que permitan su reproducción"* (Artículo 10, Real Decreto Legislativo 1/1996).

La aplicación del artículo 51 de la Ley de Propiedad Intelectual, relativo a la transmisión de los derechos del autor asalariado, adquiere una relevancia específica en el contexto de la inteligencia artificial en Recursos Humanos. En este ámbito, es cada vez más frecuente que los empleados participen en la creación de obras —como bases de datos, algoritmos, modelos predictivos o sistemas de IA aplicados a la gestión del talento— en el marco de una relación laboral.

Según el artículo 51, salvo pacto en contrario, los derechos de explotación sobre las creaciones realizadas por el trabajador en cumplimiento de sus funciones contractuales corresponden al empleador. Esto implica que, en el contexto de RR. HH., si un empleado desarrolla un algoritmo de IA o una herramienta digital para la selección de personal como parte de su trabajo, la titularidad y los derechos de explotación sobre dicha obra recaen en la empresa, y no en el trabajador individual.

Esta disposición es clave para clarificar la propiedad intelectual en proyectos de IA en RR. HH., garantizar la seguridad jurídica de las empresas y evitar conflictos sobre la titularidad de los desarrollos tecnológicos. No obstante, es recomendable que los contratos laborales y de desarrollo de IA incluyan cláusulas específicas sobre la titularidad y la posible explotación futura de los resultados, para asegurar la transparencia y el respeto a los derechos de ambas partes.

Como jurisprudencia relevante, cabe destacar la Sentencia del Tribunal Supremo, 3 de abril de 2017 (STS 172/2017), que estableció que el uso no autorizado de obras protegidas en bases de datos infringe la Ley de Propiedad Intelectual, un precedente aplicable al entrenamiento de modelos de inteligencia artificial.

2.3 Marco legal global e internacional

A nivel global, el marco legal para la implementación de IA en RR. HH. se fundamenta en la protección de los derechos fundamentales, la equidad y la inclusión, y se articula a través de normativas internacionales y directrices éticas que buscan asegurar una transformación digital responsable y justa.

El objetivo de este documento, ante la imposibilidad de abordar un análisis en detalle, es proporcionar una base rigurosa y práctica para todos los actores implicados en la aplicación, regulación y supervisión de la IA en la gestión de personas promoviendo la transparencia, la equidad y la protección de colectivos tradicionalmente excluidos.

Los países más relevantes han desarrollado marcos regulatorios y recomendaciones específicas para la IA en RR. HH., que comparten principios comunes como el consentimiento informado, la transparencia, la prevención de sesgos y la supervisión humana.

2.3.1 Organismos y marcos internacionales

La OCDE y la Unesco han publicado directrices sobre la ética de la IA, y subrayan la transparencia, la rendición de cuentas y el respeto a los derechos humanos en la toma de decisiones automatizadas (OECD *Principles on AI*, Unesco *Recommendation on the Ethics of Artificial Intelligence*). Estos principios promueven que la IA sea innovadora y confiable, y que respete los derechos humanos y los valores democráticos. La Recomendación de la Unesco sobre la Ética de la Inteligencia Artificial, aprobada en 2021, es el primer instrumento normativo global sobre la ética de la IA.

a) Principios de la OCDE sobre Inteligencia Artificial

Constituyen el primer marco internacional acordado por gobiernos para guiar el desarrollo y uso responsable de la IA. Adoptados en 2019, establecen cinco principios fundamentales, alineados con valores como la transparencia, la responsabilidad y el respeto por los derechos humanos. En el ámbito de recursos humanos, estos principios se traducen en las siguientes directrices:

1. Crecimiento inclusivo, desarrollo sostenible y bienestar. La IA debe beneficiar a todas las personas y al planeta, promoviendo el bienestar social y la inclusión. Los sistemas de IA deben diseñarse para evitar la exclusión, y garantizar oportunidades equitativas, con respeto a la diversidad y la igualdad de trato en procesos como la selección y gestión del talento.

2. Valores centrados en el ser humano y equidad. La IA tiene que respetar el estado de derecho, los derechos humanos, los valores democráticos y la diversidad. Es esencial evitar sesgos algorítmicos que puedan discriminar por género, raza, edad u otros factores protegidos, para asegurar la equidad en la toma de decisiones automatizadas.

3. Transparencia y explicabilidad. Los usuarios han de ser informados cuando interactúan con sistemas de IA, y las decisiones deben ser comprensibles y explicables. Se debe informar a candidatos y empleados sobre el uso de IA en procesos como el filtrado de currículums o la evaluación del desempeño, y proporcionar explicaciones claras sobre los criterios utilizados por los algoritmos.

4. Robustez, seguridad y protección. Los sistemas de IA deben ser seguros, confiables y resistentes a fallos o mal uso. Es necesario implementar medidas de ciberseguridad y realizar auditorías regulares para proteger los datos sensibles de empleados y candidatos, con garantía de la integridad y fiabilidad de los sistemas de IA.

5. Responsabilidad. Las organizaciones que desarrollan o implementan IA deben rendir cuentas por su impacto. Los empleadores son responsables de las decisiones tomadas por sistemas de IA, especialmente si afectan los derechos de los trabajadores, y tienen que garantizar mecanismos de revisión y reparación en caso de errores o discriminación.

b) Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial

Establece principios orientadores para un desarrollo y uso ético de la IA. Entre sus directrices más relevantes para la gestión de recursos humanos destacan:

1. **Respeto y promoción de los Derechos Humanos:** La IA debe respetar la dignidad, privacidad y derechos fundamentales de empleados y candidatos, evitando la discriminación.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

2. Inclusión, equidad y diversidad: Se destaca la necesidad de diseñar sistemas libres de sesgos, que garanticen igualdad de oportunidades y promuevan la diversidad en los procesos de selección y gestión del talento.
3. Transparencia y explicabilidad: Los procesos automatizados deben ser comprensibles y auditables, de modo que permitan a los afectados conocer los criterios y fundamentos de las decisiones algorítmicas.
4. Responsabilidad y rendición de cuentas: Las organizaciones tienen que asumir la responsabilidad por los impactos de la IA, estableciendo mecanismos efectivos de supervisión y reparación.
5. Sostenibilidad y bienestar social: La IA ha de contribuir al desarrollo sostenible y a la mejora de las condiciones laborales.

Estos principios constituyen un marco ético de referencia para la integración de IA en recursos humanos, orientando a las organizaciones hacia una gestión responsable, inclusiva y respetuosa de los derechos fundamentales (Unesco, 2021).

c) Agenda 2030 de Naciones Unidas

La Agenda 2030 de Naciones Unidas y los Objetivos de Desarrollo Sostenible (ODS 5, 8 y 10) orientan la aplicación responsable de la IA en el ámbito laboral, promoviendo la igualdad (ODS 5), el trabajo decente (ODS 8) y la reducción de desigualdades (ODS 10).

Estas normativas y directrices, aunque diversas en su alcance, comparten la finalidad de garantizar una gestión del talento justa, inclusiva y alineada con los derechos fundamentales, estableciendo obligaciones claras para empresas, desarrolladores y responsables de recursos humanos en todo el mundo.

d) Foro Económico Mundial

El Foro Económico Mundial (WEF, por su sigla en inglés) no emite normativas vinculantes, pero sí ha desarrollado directrices y principios orientativos relevantes para el uso responsable de la inteligencia artificial en recursos humanos (RR. HH.). Entre los documentos más destacados se encuentran:

• **Directrices sobre adquisiciones de IA:** El WEF ha publicado directrices para gobiernos y organizaciones sobre la adquisición y uso de sistemas de IA en áreas de alto impacto como los RR. HH. Estas directrices subrayan la necesidad de establecer reglas claras para la contratación de soluciones de IA, con atención a pruebas rigurosas, consideraciones éticas, gobernanza de datos, transparencia y supervisión humana. Se reconoce que las decisiones de IA relativas a la gestión de personas —como selección, retención, formación y beneficios— tienen consecuencias críticas para las personas y requieren escrutinio especial y controles sólidos para evitar daños y sesgos.

• **Principios éticos y manifiestos sectoriales**: El WEF impulsa la creación de principios éticos sectoriales para el uso de IA en RR. HH., y promueve la inclusión, la transparencia, la formación continua y la responsabilidad en el diseño y despliegue de estas tecnologías. Se busca que la IA complemente y potencie las capacidades humanas, pero evitar la automatización deshumanizadora y asegurar la equidad en los procesos de gestión del talento.

• **Informes sobre el futuro del trabajo:** A través de informes como *Future of Jobs Report,* el WEF

analiza el impacto de la IA en el empleo y la gestión de personas, y ya recomienda a las organizaciones adoptar marcos de gobernanza y buenas prácticas para maximizar los beneficios y mitigar los riesgos asociados a la automatización en RR. HH.

2.3.2 Estados Unidos

La supervisión recae en la Equal Employment Opportunity Commission (EEOC, 2023), organismo federal encargado de hacer cumplir las leyes antidiscriminación en el empleo, como la Civil Rights Act o la Americans with Disabilities Act. La EEOC ha emitido directrices para asegurar que los sistemas automatizados no generen prácticas discriminatorias por motivos de raza, género, edad, discapacidad u otros factores protegidos. Además, supervisa que los algoritmos utilizados en selección, evaluación y promoción no perpetúen sesgos históricos ni excluyan injustamente a determinados colectivos.

También existen leyes estatales como la California Consumer Privacy Act (CCPA 2018), que otorga a los empleados y candidatos el derecho a saber qué datos personales recopilan las empresas, cómo se utilizan y con quién se comparten; y la Illinois Biometric Information Privacy Act (BIPA 2008), que regulan el uso de datos personales y biométricos en el empleo.

El AI Bill of Rights, publicado por la Casa Blanca en 2022, establece un marco orientativo para el uso responsable de sistemas automatizados en la toma de decisiones laborales (EEOC, CCPA, BIPA, AI Bill of Rights). Incluye el derecho a la protección frente a sistemas inseguros o discriminatorios, la transparencia sobre el funcionamiento de los algoritmos, la posibilidad de intervención humana y la protección de la privacidad y los datos personales en todos los procesos automatizados.

2.3.3 Canadá

La propuesta de Artificial Intelligence and Data Act (AIDA), actualmente en proceso legislativo, exige transparencia, evaluación de riesgos sobre los sistemas automatizados y protección de derechos en el uso de IA en el empleo.

Por su parte, la Personal Information Protection and Electronic Documents Act (PIPEDA) regula el tratamiento de datos personales en el ámbito laboral (AIDA, PIPEDA). De modo análogo al RGPD, la PIPEDA exige que las empresas obtengan el consentimiento informado de los empleados y candidatos para el tratamiento de sus datos personales, y limita el uso de información sensible a los fines para los que fue recogida.

2.3.4 Brasil

La Lei Geral de Proteção de Dados (LGPD) es la ley brasileña que regula la protección de datos personales en todos los sectores, incluido el laboral. Exige que las empresas obtengan el consentimiento informado de empleados y candidatos para la recopilación, tratamiento y almacenamiento de sus datos personales, incluidos aquellos utilizados por sistemas de IA en procesos laborales.

El Marco Civil da Internet establece los principios, garantías, derechos y deberes para el uso de Internet en Brasil. Aunque su enfoque principal es la regulación del entorno digital, afecta a la gestión de RR. HH al exigir la protección de privacidad y transparencia del uso de datos en entornos digitales, incluidos aquellos tratados por sistemas de IA.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

2.3.5 Reino Unido

La United Kingdom General Data Protection Regulation (UK GDPR), que es la adaptación británica del RGPD de la Unión Europea tras el Brexit, y la Data Protection Act 2018 exigen transparencia, consentimiento y derecho a la explicación en decisiones automatizadas, mientras que la Equality Act 2010 prohíbe la discriminación en el empleo, aplicable a sistemas de IA que puedan generar sesgos (UK GDPR, Data Protection Act 2018, Equality Act 2010). Exige que los procesos automatizados sean auditables, que se minimicen los riesgos de sesgo y que se garantice la protección de los derechos de los trabajadores y candidatos.

2.3.6 China

La Personal Information Protection Law (PIPL) de China, en vigor desde 2021, es la principal norma en materia de privacidad y protección de datos personales en el país. Regula de manera estricta la recopilación, uso, almacenamiento y transferencia de datos personales, incluidos los de empleados y candidatos en procesos de RR. HH.

La Regulación sobre Algoritmos de Recomendación (2022), emitida por la Administración del Ciberespacio de China, establece requisitos específicos para los proveedores de servicios basados en algoritmos de recomendación, incluyendo aquellos utilizados en procesos de selección, evaluación y gestión de personal. Además, exige que los sistemas de IA sean auditables y que las decisiones importantes tomadas por algoritmos puedan ser revisadas por humanos.

2.3.7 Japón

Japón adopta un enfoque flexible mediante un Código de Conducta para Desarrolladores de IA, promoviendo transparencia y equidad en RR. HH:

> *"Los desarrolladores de IA deben garantizar que los sistemas sean transparentes, justos y respeten los derechos humanos"* (Código de Conducta para Desarrolladores de IA, Ministerio de Economía, Comercio e Industria de Japón, 2020).

Este código fomenta la autorregulación, pero carece de fuerza vinculante. Por ejemplo, las empresas japonesas que utilicen IA para selección deben informar a los candidatos sobre los criterios algorítmicos, pero no hay sanciones claras por incumplimiento.

2.3.8 Australia

La Privacy Act 1988 regula la protección de datos personales y la transparencia en el uso de IA en RR. HH. Otorga a los individuos derechos de acceso, corrección y eliminación de sus datos personales, y establece principios estrictos sobre seguridad, minimización y confidencialidad de la información, aplicables también a los sistemas de IA en RR. HH.

La Australian Human Rights Commission Guidelines de 2021 contiene directrices específicas sobre el uso ético y responsable de la inteligencia artificial en el ámbito laboral. Estas directrices recomiendan que los sistemas de IA sean transparentes, auditables y supervisados por humanos, y que se implementen salvaguardas para prevenir la discriminación y los sesgos algorítmicos en los procesos de selección, promoción y evaluación de personal.

2.4 Cuestiones jurídicas y derechos fundamentales en el ámbito laboral

2.4.1 Marco de cumplimiento normativo

a) Tratamiento lícito y consentimiento expreso

> *"El tratamiento de datos personales debe ser lícito, leal y transparente en relación con el interesado"* (Artículo 5.1.a, RIA 2016/679).

> *"Los datos personales solo podrán tratarse con el consentimiento del interesado, salvo en los casos previstos por la ley"* (Artículo 13, PIPL, China, 2021).

Los datos personales, como currículums, evaluaciones de desempeño o datos biométricos, deben tratarse conforme a los principios de licitud, lealtad y transparencia. Las empresas deben realizar Evaluaciones de Impacto en Protección de Datos para sistemas de IA de alto riesgo, como los que monitorean la productividad o analizan entrevistas.

b) Derecho al olvido

> *"El interesado tendrá derecho a obtener del responsable del tratamiento la supresión de los datos personales que le conciernan"* (Artículo 17, Reglamento (UE) 2016/679).

Los candidatos pueden solicitar la eliminación de sus datos tras un proceso de selección, lo que obliga a las empresas a implementar sistemas que permitan la supresión automática de datos.

c) No Discriminación

> *"Se prohíbe toda discriminación por motivos de sexo, origen racial o étnico, religión o convicciones, discapacidad, edad u orientación sexual"* (Artículo 2, Directiva 2000/78/CE).

Los algoritmos de IA deben ser diseñados y auditados para evitar sesgos que perpetúen desigualdades. Por ejemplo, un sistema de selección que favorezca a hombres sobre mujeres en roles técnicos violaría esta directiva.

d) Transparencia y explicabilidad

> *"El interesado tendrá derecho a obtener del responsable del tratamiento confirmación de si se están tratando o no datos personales que le conciernen y, en tal caso, derecho de acceso a los datos personales y a la siguiente información: a) los fines del tratamiento; b) las categorías de datos personales de que se trate; c) los destinatarios o las categorías de destinatarios a los que se comunicaron o serán comunicados los datos personales, en particular destinatarios en terceros u organizaciones internacionales; (...); g) cuando los datos personales no se hayan obtenido del interesado, cualquier información disponible sobre su origen; h) la existencia de decisiones automatizadas, incluida la elaboración de perfiles, a que se refiere el artículo 22, apartados 1 y 4, y, al menos en tales casos, información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado"* (Artículo 15, Reglamento (UE) 2016/679).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Las empresas deben explicar cómo funcionan los algoritmos de IA en RRHH, incluyendo los criterios de decisión y los datos utilizados. Esto fomenta la confianza y permite a los empleados y candidatos ejercer sus derechos. Un sistema de IA que evalúe el desempeño debe proporcionar a los empleados un informe claro sobre los factores que influyeron en su evaluación, como métricas de productividad o feedback de clientes.

e) Supervisión humana

> *"Obligaciones de transparencia para determinados sistemas de IA: Los proveedores de sistemas de IA destinados a interactuar directamente con personas físicas deberán garantizar que dichos sistemas informen claramente a los usuarios de que están interactuando con un sistema de inteligencia artificial, salvo que ello resulte evidente por el contexto; los sistemas de IA que generen contenido sintético (como imágenes, audio o texto) deberán marcar dicho contenido de manera legible, detectable y permanente como generado o manipulado por IA, a menos que se realice con fines artísticos, creativos o de sátira y sea evidente para un usuario razonable; los proveedores deberán cumplir estas obligaciones de manera proporcionada, teniendo en cuenta las capacidades técnicas y el estado de la técnica; las autoridades nacionales competentes supervisarán el cumplimiento de estas obligaciones; las excepciones a estas reglas se aplicarán en casos específicos, como cuando el sistema de IA se utilice para fines de aplicación de la ley, siempre que se cumplan las condiciones establecidas en el reglamento"* (Artículo 14, Reglamento (UE) 2024/1689).

Las decisiones críticas, como contrataciones o despidos, deben ser revisadas por profesionales capacitados, evitando la automatización total. Por ejemplo, un algoritmo que recomiende el despido de un empleado debe ser validado por un supervisor humano.

Una empresa que utilice IA para asignar promociones debe garantizar que un equipo de RRHH revise las recomendaciones del algoritmo antes de tomar decisiones finales.

2.4.2 Consideraciones prácticas para la implementación

a) Retos en la implementación a nivel global

- **Sesgos algorítmicos:** Los modelos de IA pueden perpetuar sesgos presentes en los datos de entrenamiento, como preferencias por ciertos perfiles demográficos.
- **Falta de transparencia:** Los algoritmos de caja negra dificultan la explicación de decisiones, lo que puede generar desconfianza entre empleados y candidatos.
- **Cumplimiento normativo:** Las empresas deben alinear sus sistemas con múltiples normativas (RGPD, AI Act, LOPDGDD, PIPL), lo que requiere recursos significativos.
- **Capacitación del personal:** Los equipos de RRHH necesitan formación en ética, tecnología y cumplimiento normativo para gestionar sistemas de IA.
- **Privacidad y vigilancia:** Los sistemas de monitoreo laboral basados en IA pueden vulnerar la privacidad, especialmente en contextos como China, donde la vigilancia es extensiva.
- **Costes de auditoría:** Las auditorías de sesgos y las EIPD son costosas y requieren experiencia técnica, lo que puede ser un obstáculo para las pymes.
- **Transferencias internacionales:** Las empresas multinacionales deben garantizar que las transferencias de datos cumplan con normativas como el RGPD y la PIPL, especialmente tras Schrems II (2020).

b) Recomendaciones para una implementación responsable a nivel europeo

1. Protección de datos:

o Realizar EIPD antes de implementar sistemas de inteligencia artificial de alto riesgo.
o Garantizar el derecho al olvido y limitar la retención de datos al mínimo necesario.

2. Transparencia con los empleados:

o Informar claramente a los candidatos y empleados sobre el uso de IA, incluyendo los criterios de decisión y el derecho a revisión humana.
o Publicar informes de transparencia sobre el funcionamiento de los sistemas de IA, conforme al AI Act.

3. Supervisión humana:

o Garantizar que las decisiones críticas sean revisadas por profesionales capacitados, evitando la automatización total.
o Establecer comités éticos internos para supervisar el uso de inteligencia artificial en Recursos Humanos.

4. Datos de calidad:

o Utilizar conjuntos de datos diversos y representativos para entrenar modelos de IA, minimizando el riesgo de sesgos.
o Implementar procesos de limpieza de datos para eliminar sesgos históricos.

5. Auditorías regulares:

o Implementar auditorías periódicas de los sistemas de IA para detectar y corregir sesgos, utilizando herramientas como el sandbox regulatorio español.
o Contratar auditores independientes para garantizar la objetividad.

6. Formación continua:

o Capacitar a los equipos de Recursos Humanos en ética de IA, protección de datos y cumplimiento normativo.
o Ofrecer formación técnica para comprender los principios algorítmicos.

7. Colaboración con reguladores:

o Participar en iniciativas como el Pacto por la IA y el sandbox regulatorio para garantizar el cumplimiento anticipado.
o Colaborar con la AESIA (único organismo español mencionado en estas recomendaciones) y la Oficina europea de IA para alinear los sistemas con las normativas.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

8. Mitigación de riesgos legales:

- Establecer políticas internas de cumplimiento normativo para reducir el riesgo de sanciones.
- Contratar asesores legales especializados en inteligencia artificial y protección de datos.

9. Enfoque inclusivo:

- Involucrar a colectivos infrarrepresentados en el diseño y evaluación de sistemas de IA para garantizar la equidad.
- Promover la diversidad en los equipos de desarrollo de IA.

10. Adaptación a contextos globales:

- Alinear los sistemas de IA con normativas locales, como la PIPL en China, especialmente para empresas multinacionales.
- Garantizar la compatibilidad con el RGPD en transferencias internacionales de datos.

2.5 Consideraciones éticas y legales [7]

2.5.1 Principios éticos fundamentales en el uso de la IA en RR. HH.

Implicaciones éticas de la IA en Recursos Humanos

Las disrupciones tecnológicas suelen venir acompañadas de incrementos en la eficiencia, crecimiento económico y mejoras en la calidad de vida. No obstante, también traen consigo nuevos desafíos y el riesgo de que los cambios introducidos generen efectos adversos sobre las personas y las organizaciones. La primera revolución industrial provocó una pérdida masiva de puestos de trabajo que solo fue compensada con el tiempo. Adicionalmente, la mejora en los ingresos y el nivel de vida de la población que la acompañó no fue inmediata, tomó al menos cincuenta años en consolidarse (Nardinelli, 2007). Esta transformación llegó acompañada de desafíos éticos hasta entonces imprevistos, como la explotación laboral infantil en las fábricas, condiciones de trabajo inhumanas y una creciente desigualdad (Hobsbawm, 1962, p. 40, 50). Estas condiciones prevalecieron durante décadas hasta la intervención de movimientos sociales, sindicatos, reformas legislativas y la incorporación progresiva de derechos laborales básicos.

Hoy, ante una nueva disrupción impulsada por la IA, resulta fundamental analizar los riesgos que acompañan sus promesas de eficiencia y productividad. En el ámbito de Recursos Humanos, la adopción de herramientas basadas en IA ha crecido rápidamente a nivel global, transformando

[7] **Nota**: Este subapartado completa el presente capítulo con una visión transversal. Frente a los anteriores -con un tratamiento totalmente sistematizado en torno al marco legal-, ahora se ofrece una visión conjunta desde el punto de vista ético y legal. A pesar de que es inevitable la repetición de algunas normas y regulaciones, se pretende enriquecer el contenido. El autor es Marco Tulio Daza, miembro asociado del Instituto de Ciencia de los Datos e Inteligencia Artificial (DATAI) de UNAV; ha contado con la colaboración de Esther Rodríguez-Losada, investigadora posdoctoral de DATAI. Asimismo, han aportado su visión experta dos miembros del Consejo Asesor del Proyecto IA+Igual: José Domínguez, abogado especialista en Protección de Datos y socio de Martínez-Echevarría Asociados; y Javier Moscoso, doctor en Filosofía y profesor de Investigación en el CSIC. También ha intervenido una embajadora del proyecto Estela Marín, abogada y DirCom de Sincro.

procesos clave como el reclutamiento, la evaluación de desempeño, la gestión del talento y la toma de decisiones laborales. Estas aplicaciones pueden ofrecer beneficios relevantes, como la reducción de sesgos humanos, la mejora en la eficiencia operativa y la optimización de procesos. Sin embargo, también plantean riesgos significativos que pueden provocar daños graves si no se gestionan con criterios éticos y de responsabilidad (Daza & Ilozumba, 2022).

La automatización basada en IA corre el riesgo de propiciar la toma de decisiones injustas o sesgadas (Mavrogiorgos et al., 2024), provocar el desplazamiento de empleos (Cherry, 2016), reducir la autonomía profesional, vulnerar derechos fundamentales como la privacidad y erosionar la dignidad humana (Floridi, 2016). Por ejemplo, algoritmos utilizados para filtrar currículos, analizar video-entrevistas o predecir rotación de personal perpetúan en ocasiones desigualdades preexistentes o introducen nuevas formas de discriminación inadvertida. Aunque los modelos de IA parezcan neutrales, en la práctica son capaces de favorecer o perjudicar sistemáticamente a determinados grupos demográficos (por género, origen étnico o edad, por ejemplo) debido a sesgos presentes en los datos con los que fueron entrenados o en los criterios de uso definidos por las personas que los diseñan y aplican.

Además, el riesgo no está solo en lo que la IA hace, sino en cómo lo hace: si un sistema automatizado toma decisiones sin supervisión humana, sin transparencia y sin posibilidad de apelación, puede erosionar la confianza entre la organización y sus colaboradores.

Tabla 17: *Tipología de retos éticos emergentes cuando se aplica la IA a RR. HH.*

Dimensión ética clave	Ejemplo aplicación de IA	Riesgo asociado	Necesidad de intervención
Privacidad, vigilancia y Protección de datos	Monitoreo de actividades, inferencias de personalidad, análisis emocional	Tratamiento excesivo o no consentido de datos sensibles, afectaciones a la salud física y emocional, acceso no autorizado	Minimización de datos, consentimiento informado, medidas de seguridad
Justicia y Equidad	Algoritmos de selección, promoción y despido	Reproducción de sesgos de la IA y cognitivos e inconscientes, discriminación indirecta	Auditoría algorítmica, evaluación de datos de entrenamiento y resultados
Transparencia y Rendición de cuentas	Algoritmos de selección, promoción, despido	Falta de capacidad para explicar decisiones adversas, dificultad para asignar responsabilidad, falta de transparencia en toma de decisiones automatizadas	Uso de IA explicable, consentimiento informado, documentación accesible, mecanismos de trazabilidad,
Autonomía y Dignidad	Monitoreo del comportamiento, decisiones automatizadas, asistencia basada en IA	Percepción de vigilancia y control (taylorismo digital), erosión de la confianza, pérdida de agencia, deskilling o dependencia tecnológica, riesgos a la salud.	Límites éticos, formación continua y rediseño de roles
Gobernanza	Gestión de IA	Cumplimiento regulatorio puede ser insuficiente para garantizar conducta ética	Aplicación de marcos éticos de gobernanza

Fuente: elaboración propia.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Comprender estos riesgos y anticipar sus consecuencias es esencial para garantizar que la nueva disrupción tecnológica se oriente hacia el bien común y contribuya a una transformación laboral ética, justa y sostenible.

A partir de la tipología de riesgos emergentes identificada en la literatura y evidenciada en la práctica, esta sección examina cinco dimensiones éticas clave: justicia y equidad, transparencia, responsabilidad y gobernanza, privacidad y protección de datos, y autonomía y dignidad humana. Cada dimensión plantea riesgos específicos que requieren intervenciones organizacionales concretas, tal como se resume en la siguiente tabla.

2.5.2 Privacidad, vigilancia y protección de datos

a) Impactos de la vigilancia digital sobre el bienestar laboral

Los departamentos de Recursos Humanos gestionan datos sensibles de los empleados, como información personal, datos biométricos y registros de comunicación. A ello se suman los sistemas de vigilancia laboral, que recogen información a través de equipos de trabajo, dispositivos portátiles, cámaras inteligentes, sensores y tecnologías biométricas (A. Nguyen, 2021). Estos datos alimentan sistemas de IA cada vez más integrados en distintos sectores, permitiendo generar inferencias detalladas sobre la personalidad, el estado emocional, la salud mental o incluso las creencias, la productividad y actitudes de los trabajadores. Si bien este nivel de análisis puede ser útil para ciertas decisiones organizativas, también plantea riesgos en materia de privacidad, vigilancia excesiva y puede conducir a abusos y prácticas discriminatorias.

Un estudio de la Universidad de Illinois en Chicago, que reveló que el 51 % de los empleados de almacenes de Amazon en Estados Unidos reportaron haber sufrido lesiones físicas durante su jornada laboral. El mismo informe también destaca un impacto significativo en la salud mental: el 52 % de los trabajadores se siente agotado y el 35 % afirma que su empleo tiene un efecto negativo en su bienestar psicológico (Gutelius & Pinto, 2023). A esto se suman los hallazgos de las autoridades de seguridad laboral del estado de Washington, que han identificado una relación directa entre los sistemas de monitoreo de algunas compañías y el aumento de trastornos musculoesqueléticos entre sus empleados (Day, 2023).

Tecnologías como el análisis emocional en entrevistas por vídeo, el monitoreo de comunicaciones internas y los asistentes conversacionales inteligentes han ampliado significativamente el alcance de la vigilancia corporativa, muchas veces sin que los empleados comprendan plenamente su funcionamiento ni hayan otorgado un consentimiento explícito. Esto plantea riesgos de tratamiento excesivo o injustificado de datos personales, especialmente cuando se utilizan para elaborar perfiles conductuales o predecir comportamientos futuros. Prácticas como el talent analytics combinan datos de desempeño, historial médico y patrones de comunicación para anticipar riesgos de rotación o caídas en la productividad (Qin et al., 2023), lo que puede derivar en decisiones discriminatorias o injustas. Adicionalmente existe el riesgo de acceso no autorizado a los datos personales. Un ejemplo que ilustra este riesgo es el caso de Ashley Madison, donde una filtración expuso información íntima de millones de usuarios; muestra el grave impacto de una gestión inadecuada de datos sensibles (Mansfield-Devine, 2015).

Para mitigar estos riesgos, se han implementado diversos marcos regulatorios y normativas orientadas a proteger la privacidad y los datos personales.

b) Regulaciones sobre privacidad de datos

En este contexto, las regulaciones sobre privacidad de datos cumplen una doble función: actúan como límites legales vinculantes, pero también como garantías éticas para la protección de derechos fundamentales. En Europa, el Reglamento General de Protección de Datos (RGPD) representa el marco normativo más robusto en esta materia, estableciendo principios como la limitación de la finalidad, la minimización del uso de datos, la obligación de transparencia y el derecho a no ser objeto de decisiones automatizadas sin supervisión humana (Reglamento (UE) 2016/679).

Estos principios se refuerzan a través de normativas nacionales —como la Ley Orgánica 3/2018 en España— que introducen salvaguardas adicionales en el ámbito laboral (BOE, 2018). Por ejemplo, se exige que la evaluación automatizada del rendimiento laboral sea proporcional, justificada y comunicada previamente a la persona afectada, preservando en todo momento su derecho a impugnar la decisión y obtener una intervención humana significativa (AEPD, 2022).

Desde una óptica ética, estas regulaciones no deben entenderse como una barrera al despliegue de la IA, sino como una oportunidad para rediseñar los sistemas desde una lógica de respeto y confianza. La protección de datos personales deja de ser un mero requisito técnico para convertirse en un criterio de calidad y legitimidad del sistema. Las organizaciones que aspiran a una implementación ética de la IA deben incorporar, desde la fase de diseño, enfoques proactivos como la privacidad desde el diseño y por defecto (EDPB, 2021).

Además, propuestas legislativas como el RIA refuerzan la necesidad de coordinación entre la normativa de protección de datos y la regulación del uso de sistemas algorítmicos, especialmente en contextos laborales de alto riesgo (European Commission, 2021). Esta convergencia normativa plantea un nuevo estándar: no basta con cumplir con las leyes de protección de datos, también es necesario demostrar que los sistemas de IA son auditables, justos y comprensibles (Wachter et al., 2017).

2.5.3 Justicia, equidad y rendición de cuentas

a) Justicia y equidad: sesgo de la IA

Uno de los principales riesgos éticos del uso de IA en Recursos Humanos es la posibilidad de que los modelos generen decisiones parciales o discriminatorias por sesgos incorporados en los sistemas o en los propios operadores humanos. El sesgo en la IA se refiere a errores sistemáticos en los resultados de los modelos que producen discriminación o favoritismo hacia ciertos grupos (Berma & Rubin, 2018). Tal es el caso de los sistemas de reconocimiento facial, que pueden mostrar un rendimiento superior con ciertos perfiles demográficos, o del software de reclutamiento, que puede desfavorecer a un grupo específico. Además, el uso de sistemas de IA, a menudo percibidos como objetivos y precisos, es capaz, también, de detonar sesgos cognitivos e inconscientes en quienes los utilizan. Por ejemplo, una confianza excesiva en los resultados de la IA puede derivar en decisiones injustas y debilitar la capacidad de juicio crítico de los usuarios.

El sesgo en la IA puede tener múltiples orígenes, entre ellos: desigualdades históricas, estereotipos y prejuicios sociales presentes en los datos de entrenamiento; errores en el diseño de los modelos; la propia naturaleza de los algoritmos de aprendizaje automático; y decisiones humanas defectuosas durante el desarrollo o implementación del sistema (Neroli et al., 2021; Tarik et al., 2019).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

La imparcialidad absoluta es extremadamente difícil de lograr, si no imposible, incluso en sistemas automatizados. Elegir o tomar una decisión implica discriminar entre diferentes opciones a partir de criterios establecidos. El problema surge porque distintas personas, y culturas organizacionales, valoran de manera diferente qué criterios son más relevantes, lo que conduce inevitablemente a favorecer determinadas concepciones de "éxito" o "mérito" sobre otras. Considera un algoritmo de contratación que podría asignar mayor peso a la trayectoria académica que a la experiencia práctica, introduciendo así un sesgo valorativo que no necesariamente refleja las capacidades reales para el puesto.

Además, lo que se considera moralmente aceptable no es estático, sino que evoluciona constantemente. Antes del siglo XIX, el trabajo infantil era socialmente aceptado e incluso visto como necesario para la economía familiar. Del mismo modo, antes del siglo XX, el derecho al voto de las mujeres no era reconocido en la mayoría de los países. Las normas, expectativas y sensibilidades cambian con el tiempo y varían según el contexto cultural y organizacional. Esto tiene implicaciones directas en los modelos de IA que son entrenados con datos históricos, pues los hacen propensos a reproducir y amplificar desigualdades estructurales. Aunque las herramientas de IA prometen eliminar la parcialidad humana, en realidad su objetividad depende por completo de los datos con los que se entrenan y de los criterios de diseño incorporados por los equipos técnicos.

Comprender cómo surgen los sesgos en la IA es esencial para abordarlos de manera efectiva. La literatura académica ofrece diversas clasificaciones, como la distinción entre sesgos estadísticos y sociales (Dong et al., 2022; Narayanan, 2018; Neroli et al., 2021), o entre sesgos técnicos y aquellos relacionados con la interacción entre humanos y sistemas de IA (Hendricks, 2024).

No obstante, en este análisis adoptamos una tipología ampliamente reconocida que identifica tres fuentes principales de sesgo: los datos utilizados para entrenar el modelo, el propio diseño y funcionamiento del algoritmo, y la forma en que el sistema es implementado o utilizado por personas (Barocas et al., 2023; Bellamy et al., 2019; Ferrara, 2023; Mehrabi et al., 2019). Estas tres dimensiones —datos, modelo e implementación— son especialmente relevantes en el ámbito de los Recursos Humanos, donde la toma de decisiones automatizadas puede tener un impacto directo en los derechos de los trabajadores.

Sesgos a nivel de datos

La frase "basado en datos" (data bias) suele asociarse con objetividad, pero si los datos están sesgados, sus resultados también lo estarán. Los algoritmos de aprendizaje automático son entrenados con grandes volúmenes de datos los cuales pueden contener sesgos y prejuicios humanos.

Un caso emblemático es el de un software de reclutamiento basado en inteligencia artificial, que fue entrenado con currículos recopilados por la empresa durante una década. Aunque el algoritmo no utilizaba explícitamente información sobre el género para filtrar candidaturas, aprendió a penalizar los currículos que contenían términos asociados con mujeres, como referencias a estudios en colegios femeninos.

La causa residía en que la mayoría de las contrataciones previas realizadas por la organización habían sido hombres, reflejando una histórica predominancia masculina en el sector tecnológico (Chang, 2023). Como resultado, los datos de entrenamiento contenían un sesgo de género que el modelo reprodujo, dando lugar a una forma de discriminación indirecta sistemática.

En procesos de promoción o evaluación, si los datos de entrenamiento contienen prejuicios (evaluaciones subjetivas influenciadas por estereotipos), un sistema de inteligencia artificial que use esos registros puede terminar otorgando sistemáticamente calificaciones más bajas a ciertos grupos.

Asimismo, la falta de representación de minorías en los datos de entrenamiento puede generar errores al evaluar perfiles que se desvían del patrón dominante. Un algoritmo entrenado principalmente con datos de una región o cultura específica podría no interpretar adecuadamente las características de candidatos provenientes de contextos distintos, un fenómeno conocido como sobreajuste a un entorno particular (overfitting) (Barocas et al., 2023). Así, una IA entrenada con entrevistas en inglés estadounidense podría malinterpretar la entonación, el lenguaje corporal o los patrones comunicativos de candidatos de otras culturas, calificándolos erróneamente como inseguros o poco expresivos.

Sesgos a nivel del modelo

El sesgo algorítmico (Algorithmic model bias) surge durante las etapas de diseño, entrenamiento y validación en el desarrollo de modelos de IA. Es causado por la propia naturaleza de las técnicas de aprendizaje automático y por decisiones tomadas por ingenieros y científicos de datos, como la definición de funciones, la arquitectura del modelo o la selección de variables, que pueden favorecer inadvertidamente ciertos resultados o grupos demográficos en detrimento de otros.

En un caso paradigmático, un algoritmo de IA asignó el mismo nivel de riesgo a pacientes afroamericanos gravemente enfermos que a pacientes blancos menos enfermos (Norori et al., 2021), lo que redujo a la mitad el número de pacientes afroamericanos identificados para recibir atención adicional. Esto es debido a que el algoritmo fue entrenado para usar el gasto sanitario como variable sustituta (o proxy) de las necesidades de salud, asumiendo erróneamente que los pacientes afroamericanos estaban más sanos que los pacientes blancos con la misma gravedad clínica, simplemente porque históricamente se gastaba menos dinero en ellos (Obermeyer et al., 2019).

Otra fuente de sesgo derivada de decisiones técnicas puede ser fijar umbrales o parámetros (por ejemplo, en el reclutamiento de personal) sin evaluar cómo afectan a distintos grupos, o usar categorías demasiado simplificadas que no captan la complejidad del comportamiento humano. Además, los desarrolladores pueden priorizar resultados que confirmen sus expectativas, ignorando señales de bajo rendimiento para ciertos subgrupos. Ilustrativamente, un algoritmo de evaluación de desempeño podría funcionar bien para empleados de departamentos técnicos, pero ser impreciso con perfiles creativos o administrativos.

Por otra parte, la propia lógica de los algoritmos de aprendizaje automático favorece que los modelos de IA infieran información sensible a través de variables sustitutas, incluso cuando atributos como el género, preferencia sexual o religión han sido excluidos deliberadamente del entrenamiento. Este enfoque, conocido como fairness through unawareness, parte de la premisa de que eliminar las variables sensibles previene la discriminación, pero en la práctica no garantiza decisiones equitativas. El caso del software de reclutamiento descrito anteriormente ilustra este problema, aunque el género fue eliminado de los datos a analizar, el sistema penalizaba currículos que contenían términos asociados con mujeres.

Este fenómeno ocurre porque muchas variables aparentemente neutrales están correlacionadas con atributos sensibles. Como muestra, el código postal puede actuar como indicador del nivel socioeconómico o del origen étnico en contextos donde existe segregación urbana. De forma similar,

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

un algoritmo puede deducir el género (y otros atributos) de un candidato a partir de su nombre, sus aficiones o incluso su estilo de escritura (Qin et al., 2023), reproduciendo sesgos discriminatorios a pesar de haber "eliminado" la variable protegida. En consecuencia, confiar únicamente en la ausencia de datos sensibles no es suficiente para asegurar decisiones justas y equitativas.

Sesgos a nivel de despliegue o uso

Que un algoritmo funcione correctamente en un entorno controlado no garantiza que no aparezcan nuevos sesgos en su aplicación real. En los sistemas sociotécnicos, la interacción con usuarios humanos y entornos dinámicos puede influir tanto en las decisiones humanas como en el propio comportamiento del modelo, que tiende a adaptarse o evolucionar con el uso. Es en este contexto donde surge el sesgo de despliegue (deployment bias), más sutil y difícil de detectar, pero con importantes implicaciones éticas y prácticas.

Esto ocurre, por ejemplo, a través de bucles de retroalimentación (feedback loops). Las predicciones de la IA influyen en el comportamiento de las personas, y esas reacciones alimentan nuevamente al sistema, cerrando un ciclo (Hendrycks, 2024). Si un algoritmo de evaluación de desempeño califica más bajo a cierto colectivo (una razón posible es el uso de datos de entrenamiento sesgados), los supervisores, influidos por esa etiqueta bajo desempeño, podrían dar menos oportunidades de desarrollo a esas personas, amplificando el sesgo inicial y dando lugar a una “profecía autocumplida”. Luego los datos de desempeño empeorados por esta decisión retornan al sistema reforzando la asociación negativa.

Otra faceta del sesgo de despliegue es el riesgo de que aparezcan sesgos inconscientes o sesgos cognitivo-humanos. Muchas veces, la sola presencia de una recomendación algorítmica puede alterar la decisión final del humano. Un ejemplo común es el “sesgo de automatización” (automation bias), por el que los usuarios tienden a ceder a las sugerencias del algoritmo sin someterlas a suficiente escrutinio crítico, asumiendo que la máquina es objetiva y precisa (ForHumanity, 2023). Un reclutador podría dar por buena la puntuación de un candidato arrojada por la IA y omitir realizar una evaluación propia más exhaustiva, confiando ciegamente en el sistema. Este exceso de confianza puede hacer que se pasen por alto errores obvios o consideraciones contextuales importantes, llevando a decisiones injustas o subóptimas.

Otros ejemplos de sesgos cognitivos e inconscientes susceptibles de amplificarse con el uso de IA incluyen el sesgo de confirmación (confirmation bias), en el que el usuario interpreta los resultados del sistema como evidencia que refuerza sus propias creencias previas, y el efecto de anclaje (anchoring effect), donde la primera puntuación o recomendación proporcionada por el modelo condiciona desproporcionadamente el juicio posterior (Tversky & Kahneman, 1974). Como advierte Kahneman (2011), el pensamiento humano opera a menudo mediante un sistema rápido e intuitivo (“Sistema 1”) que busca atajos mentales para ahorrar esfuerzo, lo cual facilita que las decisiones algorítmicas se acepten sin el análisis pausado y reflexivo del “Sistema 2”, más lento y deliberativo. Esto no solo reduce la calidad del juicio, sino que puede desplazar la responsabilidad individual al sistema, debilitando la rendición de cuentas.

b) Transparencia y la rendición de cuentas

La transparencia en IA aplicada a Recursos Humanos abarca varias dimensiones. Primero, la transparencia hacia los candidatos y empleados: las personas tienen derecho a saber cuándo una

decisión que les afecta (ser preseleccionado para un empleo o recibir una determinada evaluación de desempeño) ha sido influenciada por un algoritmo. Esto se debe a que las decisiones automatizadas pueden tener un impacto significativo en sus oportunidades laborales, y conocer el grado de intervención algorítmica es necesario para garantizar la posibilidad de impugnar decisiones injustas. Adicionalmente, no basta con informar que se ha utilizado IA en una decisión; es necesario poder explicar, en un lenguaje claro y comprensible, los factores principales que influyeron en ese resultado. El problema es que algunos modelos de IA, como las redes neuronales profundas, funcionan como una especie de caja negra: procesan enormes cantidades de datos a través de múltiples pasos internos, pero no ofrecen una explicación sencilla de cómo llegaron a una conclusión. Esto hace muy difícil, incluso para los expertos, entender o justificar por qué se tomó una decisión.

Debido a esto, la explicabilidad (explicability), entendida como la integración tanto de la inteligibilidad como de la rendición de cuentas (accountability), es un elemento esencial de la transparencia (Floridi et al., 2018). Sin ella, los posibles sesgos del sistema permanecen ocultos y las personas afectadas carecen de medios para comprender, cuestionar o impugnar los resultados. En 2021, un tribunal en Países Bajos revirtió el despido de varios conductores de Uber porque la empresa no pudo explicar adecuadamente la decisión, que había sido tomada exclusivamente por un algoritmo. El juez dictaminó que despedir "basado únicamente en procesamiento automatizado" vulneraba la normativa (invocando el Art. 22 del GDPR), y ordenó la reincorporación de los trabajadores (Nawrat, 2021).

Por otra parte, la rendición de cuentas exige que las organizaciones designen claramente quién se hace responsable de las decisiones asistidas por IA. El riesgo es que, ante un error, por ejemplo, un despido injustificado o una recomendación errónea, ninguna persona o área pueda asumir la responsabilidad efectiva. Este problema se agrava con el aumento de la autonomía de los algoritmos, que dificulta rastrear cómo se tomó una decisión y quién debe responder por sus consecuencias. Cuando una decisión automatizada causa daño, aparecen serios desafíos para asignar la responsabilidad entre las partes involucradas dado que la culpa solo puede atribuirse justificadamente cuando el actor tiene cierto grado de control (Matthias, 2004) e intencionalidad (Mittelstadt et al., 2016).

El modelo tradicional de rendición de cuentas resulta insuficiente en estos casos, ya que ningún actor individual parece tener un control total sobre el comportamiento del algoritmo una vez desplegado (Mittelstadt et al., 2016). Las empresas que desarrollan estas tecnologías pueden alegar que la responsabilidad recae en los ingenieros que diseñaron los modelos, mientras que los desarrolladores pueden argumentar que su función fue meramente técnica y que no decidieron sobre su uso final. A su vez, aún y cuando el sistema haya sido diseñado y desplegado correctamente, las personas que lo utilizan pueden abusar de él o aplicarlo de forma negligente. Un caso ilustrativo podría ser un gerente que usa un sistema de evaluación de desempeño para justificar el despido de un empleado incómodo, manipulando previamente los datos que alimentan al algoritmo.

Para mitigar este tipo de riesgos, dos estrategias son clave: por un lado, incorporar el juicio humano (human-in-the-loop) en aquellas decisiones automatizadas que tienen un impacto directo sobre las personas (Ozmen Garibay et al., 2023); por otro, garantizar la trazabilidad (traceability) de los procesos algorítmicos, de modo que se pueda comprender y reconstruir cómo se llegó a una decisión (Mittelstadt et al., 2016).

No todas las tareas requieren este nivel de supervisión: algunas pueden automatizarse de manera segura, como la generación de métricas o el análisis del clima organizacional, siempre que no se

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

realicen inferencias individualizadas. En cambio, en procesos más sensibles como la selección de personal, la intervención humana es indispensable. Aunque un sistema de IA pueda filtrar y puntuar automáticamente a candidatos para un puesto, la decisión final sobre a quién entrevistar o contratar debe ser revisada y validada por un profesional de Recursos Humanos. Esto permite tener en cuenta factores contextuales, trayectorias no convencionales o aspectos cualitativos que el sistema podría no interpretar adecuadamente.

La trazabilidad implica documentar de forma completa y comprensible el recorrido que sigue un sistema de IA desde los datos iniciales hasta las decisiones finales. Esto permite entender cómo se generan los resultados, auditar el proceso y corregir posibles errores o sesgos. Para que esta trazabilidad sea efectiva, es clave mantener un registro tanto de las recomendaciones generadas por la IA como de las decisiones humanas que se tomen a partir de ellas, de modo que el proceso pueda reconstruirse y evaluarse posteriormente. En una implementación básica, debe incluirse al menos el origen y las características de los datos utilizados, las transformaciones aplicadas, los parámetros del modelo, los criterios de evaluación y las acciones derivadas de los resultados (Mora-Cantallops et al., 2021).

Este registro facilita la práctica de auditorías algorítmicas periódicas, donde terceros o equipos internos independientes examinan el sistema en busca de comportamientos discriminatorios o problemas de rendimiento. Algunas empresas están sometiendo sus herramientas a auditorías algorítmicas. La empresa HireVue encargó una auditoría externa a O'Neil Risk Consulting para evaluar posibles sesgos en su software de entrevistas por IA, y publicó un resumen con resultados positivos y recomendaciones.

Las empresas que priorizan una utilización de la IA ética y transparente proyectan una imagen positiva y construyen confianza, tanto internamente como ante candidatos externos. En contraste, una cultura opaca, basada en 'decisiones secretas', puede generar desconfianza e incidir negativamente en la moral.

2.5.4 Autonomía y dignidad

Los problemas analizados en las secciones previas, como las violaciones a la privacidad, la vigilancia intensiva, los sesgos algorítmicos e inconscientes humanos, la falta de transparencia y la ausencia de rendición de cuentas, convergen todos en este punto: la amenaza de causar daños que socavan la autonomía y la dignidad de las personas. Esto ocurre debido a dos causas fundamentales: la posible utilización de los propios datos de los individuos en su contra, y la delegación, o al menos la influencia significativa, de decisiones relevantes a sistemas automatizados que pueden incorporar sesgos, ofrecer resultados erróneos o ignorar las particularidades de cada caso. Esta erosión no es necesariamente consecuencia de acciones deliberadas con la intención de causar daño, sino el resultado acumulativo de múltiples prácticas tecnológicas que, al deshumanizar la toma de decisiones, debilitan la capacidad de las personas para actuar libremente. De esta forma, las personas corren el riesgo de ser instrumentalizadas para alcanzar fines organizacionales, en lugar de ser reconocidas como fines en sí mismas.

La autonomía se ve comprometida cuando decisiones fundamentales sobre la trayectoria laboral, como la contratación, la promoción o el despido, se delegan a sistemas automatizados opacos, muchas veces sin posibilidad de revisión, explicación o apelación. Estas decisiones, cuando son mediadas por algoritmos, no permiten el diálogo ni explicitan sus criterios y pueden reducir al individuo a un objeto

pasivo dentro del proceso. La confianza en la supuesta objetividad de los sistemas de IA, conocida como el "halo algorítmico", refuerza la renuncia a una evaluación ética activa, al fomentar una aceptación acrítica de los resultados y desplazar el juicio profesional.

La privacidad es un componente central de la autonomía porque permite a las personas gestionar diferentes tipos de relaciones sociales (Rachels, 1975) al controlar que información comparten y con quien. Cuando esta capacidad se ve restringida por prácticas como la recolección masiva de datos, la monitorización constante o el análisis emocional, la persona pierde control sobre la forma en que es percibida, evaluada y representada. La vigilancia continua, muchas veces descrita como una forma de Taylorismo digital (Cole et al., 2022), genera una sensación de observación permanente que erosiona la confianza y fomenta la autocensura.

La dignidad, por su parte, se ve vulnerada cuando los datos personales del trabajador, recopilados mediante sistemas de monitoreo, análisis emocional o evaluaciones digitales, son utilizados para emitir inferencias que pueden ser invasivas, imprecisas o perjudiciales. El hecho de que una organización utilice sistemas que pueden ofrecer resultados sesgados e imprecisos para predecir comportamientos, actitudes o niveles de compromiso, convierte al individuo en un recurso instrumental antes que en un sujeto de derechos. Además, el uso de los datos personales en contra de los propios individuos, instrumentalizándolos sin una justificación legítima ni un consentimiento claro, niega el reconocimiento pleno de su individualidad y de su valor moral inherente (Daza-Ramírez et al., 2025).

A ello se suma el fenómeno del deskilling, o pérdida progresiva de habilidades humanas, que ocurre cuando la IA sustituye competencias clave como el juicio contextual, la empatía o la deliberación crítica (Sison et al., 2023). La confianza excesiva y la dependencia tecnológica debilitan el desarrollo profesional y la capacidad de análisis crítico (Daza-Ramirez & Luna Rizo, 2025) al mismo tiempo, dando lugar a una lógica organizativa en la que la intervención humana se limita a validar decisiones ya tomadas por la máquina. Combinado con otros sesgos inconscientes que estos sistemas pueden activar, este fenómeno no solo compromete la calidad de las decisiones, sino también el sentido de agencia y responsabilidad personal que fundamenta la autonomía de quienes participan en el proceso.

Frente a estos riesgos, es necesario establecer límites éticos que garanticen que las personas sigan siendo el centro de las decisiones laborales. La supervisión humana significativa en procesos automatizados, el rediseño de roles para que la IA complemente, y no sustituya, las capacidades humanas, y la formación en literacidad de la IA que permita a los trabajadores interactuar de manera crítica con estas tecnologías, son condiciones necesarias para preservar la autonomía y proteger la dignidad. Para esto, existen distintos marcos y directrices que buscan ofrecer orientación sobre la gobernanza de la IA en el entorno laboral. A continuación, examinaremos algunos de los más importantes.

2.5.5 Gobernanza y toma de decisiones éticas

a) Gobernanza: marcos y directrices éticas para la IA

La adopción de IA en el ámbito de Recursos Humanos no puede limitarse solamente a consideraciones técnicas ni al mero cumplimiento legal. Esto se debe a que, en muchos casos, la tecnología avanza más rápido que las regulaciones, y en otros, las normas pueden presentar vacíos legales o no contemplar todos los escenarios posibles.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Dado el impacto que estos sistemas pueden tener en decisiones laborales críticas, como la contratación, la promoción, el despido o la evaluación del desempeño, resulta indispensable contar con una gobernanza ética sólida que defina principios orientadores claros y establezca mecanismos internos eficaces para llevarlos a la práctica. En los últimos años, diversos organismos internacionales, centros de investigación y redes profesionales han desarrollado marcos éticos con ese propósito: asegurar que la implementación de la IA se realice de forma justa, transparente, trazable y respetuosa de la dignidad humana. Estos marcos promueven un enfoque de autorregulación basado en el compromiso voluntario de las organizaciones con la integridad.

Entre los marcos más influyentes destaca, en primer lugar, el Marco de Gestión de Riesgos de IA desarrollado por el Instituto Nacional de Estándares y Tecnología de Estados Unidos (NIST, 2023). Este documento ofrece una guía detallada para identificar y gestionar los riesgos éticos y técnicos asociados con los sistemas de IA a lo largo de todo su ciclo de vida. A diferencia de otros enfoques más normativos, el marco del NIST se centra en ayudar a las organizaciones a construir sistemas que sean, desde el diseño, explicables, justos, seguros, resilientes, transparentes y respetuosos de la privacidad.

En el contexto europeo, la Comisión Europea formuló en 2019 las Directrices Éticas para una IA Fiable (Comisión Europea, 2019) y han sido particularmente influyentes. Estas directrices fueron desarrolladas por un Grupo de expertos de alto nivel y establecen siete requisitos clave para una IA confiable: supervisión humana, solidez técnica y seguridad, privacidad y gestión de datos, transparencia, diversidad, no discriminación y equidad, bienestar social y medioambiental, y rendición de cuentas. Su énfasis en la dimensión social y su aplicabilidad transversal las han convertido en un referente para muchas organizaciones que buscan integrar principios éticos en sus procesos de digitalización.

A nivel global, la Recomendación sobre la Ética de la IA de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco, 2021) constituye otro importante marco ético para la gobernanza de la IA. Aprobada por 193 países, esta recomendación destaca por su enfoque basado en derechos humanos, inclusión y sostenibilidad. Reconoce que la IA no es tecnológicamente neutral y llama a los Estados y a las organizaciones a adoptar marcos institucionales que aseguren su uso responsable.

Los Principios de la OCDE sobre IA (OCDE, 2019) adoptados por los países miembros y respaldados por el G20. Este marco establece cinco principios orientadores para un desarrollo y uso responsable de la IA: crecimiento inclusivo y bienestar sostenible, respeto a los derechos humanos y valores democráticos, transparencia y explicabilidad, solidez técnica y seguridad, y rendición de cuentas. Su enfoque destaca la necesidad de que los sistemas de IA no solo sean técnicamente eficaces, sino también socialmente beneficiosos. Estos principios han influido en múltiples estrategias nacionales de IA y ofrecen una base sólida para que las organizaciones de Recursos Humanos adopten tecnologías con una visión equilibrada entre eficiencia y responsabilidad social.

Otro referente clave en la gobernanza ética de la IA, específicamente en el ámbito de los Recursos Humanos, es el Toolkit Human-Centred AI for HR del Foro Económico Mundial (WEF, 2021), que ofrece una guía práctica para implementar inteligencia artificial en Recursos Humanos de forma ética y centrada en las personas. Desarrollado en colaboración con empresas, expertos y organizaciones sociales, el toolkit aborda temas clave como la gestión del sesgo algorítmico, la protección de datos de empleados, la explicabilidad de los sistemas y la supervisión humana en decisiones automatizadas. Su

enfoque se basa en integrar principios éticos desde la planificación y adquisición de tecnologías hasta su evaluación y monitoreo continuo. Destaca por traducir valores como la equidad, la transparencia y la rendición de cuentas en herramientas aplicables a los procesos de selección, evaluación y gestión del talento.

Junto a estos marcos ampliamente reconocidos, existen otras iniciativas relevantes que complementan el panorama internacional. El Ada Lovelace Institute ha promovido principios de justicia algorítmica, explicabilidad y participación democrática en el diseño de sistemas (Ada Lovcelace, 2022). El IEEE, a través del marco Ethically Aligned Design, ha integrado consideraciones éticas directamente en los procesos de ingeniería de IA. La Declaración de Montreal (Universidad de Montreal, 2024) ha propuesto principios como la solidaridad, la participación y el desarrollo sostenible. Y los *Principios de Asilomar* (Future of Life Institute, 2017) ofrecen orientaciones sobre el uso seguro y alineado de la IA con los valores humanos.

Sin embargo, la existencia de estos marcos éticos no garantiza por sí sola su aplicación efectiva. Uno de los principales retos de la gobernanza es su "operacionalización", es decir, traducir principios abstractos en políticas, procedimientos y estructuras organizativas concretas. La instrumentación de estrategias al interior de las organizaciones como comités interdisciplinarios de supervisión, implementar auditorías algorítmicas periódicas, reforzar los mecanismos de trazabilidad, y promover una cultura interna orientada a la rendición de cuentas. Para esto, es necesario adoptar un enfoque integral que vaya más allá del cumplimiento normativo, y que refuerce los enfoques de autorregulación e integridad organizacional. Esto implica promover el desarrollo del carácter moral de los integrantes de la organización, de modo que estén mejor preparados para afrontar dilemas éticos cuando surjan. Solo mediante este tipo de estructuras es posible garantizar que la IA opere como una herramienta al servicio del desarrollo humano. El siguiente apartado aborda cómo podría llevarse a la práctica este enfoque.

b) Cómo afrontar los dilemas éticos

Ante los escenarios planteados, resulta evidente que las organizaciones deben estar preparados para identificar y abordar los dilemas éticos que emergen del uso de IA en la gestión del talento. Preguntas como "¿debo confiar en la recomendación del algoritmo, aunque intuya que perjudica a ciertos grupos?" o "¿hasta qué punto conviene automatizar un proceso a costa de reducir la intervención humana?" ilustran los desafíos cotidianos que enfrentan quienes trabajan en este ámbito. No existen respuestas simples ni universales, pero sí enfoques estratégicos y herramientas que permiten afrontar estos dilemas de forma estructurada y fundamentada en valores.

Una primera línea de defensa ética frente a los dilemas que plantea la IA en Recursos Humanos es el establecimiento de políticas internas claras sobre su uso. Así como existen manuales de cumplimiento normativo, deberían definirse protocolos específicos que indiquen qué prácticas están permitidas, cuáles se prohíben expresamente; tal es el caso del uso de IA sin supervisión humana en decisiones críticas, o cómo proceder ante situaciones de ambigüedad. Estas políticas deben contemplar escenarios previsibles y prever mecanismos de doble verificación en decisiones automatizadas relevantes, como contrataciones, promociones o despidos. A su vez, es recomendable apoyarse en marcos de referencia y comunidades expertas externas como las analizadas en la sección anterior.

Ante la posibilidad de que la IA ofrezca resultados sesgados, imprecisos o erróneos, la supervisión humana también resulta indispensable en decisiones críticas. Mantener a personas "en el circuito"

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

(human in the loop) no solo mejora la calidad de las decisiones, sino que facilita la rendición de cuentas (Ozmen Garibay et al., 2023). La participación de empleados y candidatos en el ecosistema algorítmico también es clave. Habilitar canales de retroalimentación, realizar encuestas sobre la percepción de justicia en procesos automatizados y ofrecer mecanismos de apelación son anticipar problemas. Escuchar las voces de quienes viven las consecuencias de estas decisiones puede revelar problemas invisibles desde la dirección.

Otra estrategia de anticiparse a los problemas antes de que ocurran es integrar principios éticos desde el diseño del sistema, lo que se conoce como ethics by design (Brey & Dainow, 2023). En paralelo, se recomienda realizar auditorías algorítmicas y establecer sistemas de monitoreo continuo. Incluir evaluaciones de impacto ético en las primeras fases del desarrollo, aplicar métricas de equidad junto con las de eficiencia y optar por modelos más interpretables son ejemplos de buenas prácticas. Si un sistema de evaluación está programado desde el inicio para no superar ciertos umbrales de disparidad entre grupos o se establecen criterios de desempeño adecuados, se evita tener que corregirlo tras detectar un daño.

Sin embargo, el establecimiento de estos parámetros no es sencillo. Si se considera que el 70.4 % de los jugadores de la NBA de la temporada 2023 eran afroamericanos (Statista, 2023). ¿Tendría sentido cambiar esa composición en aras de la diversidad? Claramente no, en este caso la distribución es resultado exclusivamente del mérito deportivo. Esto es diferente en otros sectores como la salud o en la asignación de beneficios donde el foco ético recae en garantizar un acceso equitativo para todos los grupos. Por ello, es fundamental contar con profesionales capacitados y con criterio para el despliegue y operación de sistemas de inteligencia artificial en Recursos Humanos. No basta con que el personal a cargo tenga dominio de las funciones administrativas y un conocimiento profundo de la industria; también es indispensable una formación sólida en literacidad digital y ética de la IA. Según un estudio de General Assembly, el 58 % de los líderes empresariales en Estados Unidos y Reino Unido no ha recibido formación en IA (Business Wire, 2024), lo que revela una brecha significativa que incluye aspectos técnicos y éticos. Reducir esta brecha requiere implementar talleres, elaborar guías internas y fomentar espacios de reflexión dentro de las organizaciones.

Otra dimensión crítica es la transparencia técnica. Los sistemas de IA deben ser comprensibles al menos en sus aspectos fundamentales por quienes los usan. Si un responsable de RR. HH. no puede explicar por qué el algoritmo seleccionó a un candidato, difícilmente podrá justificar o revisar esa decisión en caso de conflicto. Apostar por herramientas explicables, interfaces accesibles y reportes interpretables refuerza tanto la confianza como la rendición de cuentas.

A nivel filosófico, la ética de la virtud ofrece una guía valiosa para enfrentar dilemas en los que no existe una única respuesta correcta. En lugar de limitarse a la pregunta “¿está permitido?”, este enfoque propone una reflexión más profunda: “¿es lo correcto?”, considerando no solo los resultados, sino también la intención y el carácter moral de quien toma la decisión. Virtudes como la justicia, la prudencia y la honestidad orientan la acción más allá del cumplimiento normativo o de los códigos externos. En contextos donde la regulación es insuficiente y donde la autorregulación o la integridad corporativa pueden ser instrumentalizadas para proyectar una imagen ética sin un compromiso real, el cultivo del carácter moral al interior de las organizaciones puede proporcionar una motivación intrínseca para actuar de manera ética. Así, la ética de la virtud no sustituye, sino que complementa y refuerza los enfoques normativos y procedimentales, dotando de sustancia moral a las decisiones en torno al uso de la IA.

Enfrentar los dilemas éticos que plantea la IA en Recursos Humanos requiere un equilibrio entre principios y pragmatismo. La tecnología, bien utilizada, puede reforzar la equidad, corregir sesgos humanos y aumentar la objetividad. Pero mal implementada puede amplificar injusticias, diluir responsabilidades y erosionar la confianza organizacional.

2.5.6 Aspectos legales y cumplimiento normativo

a) Aspectos legales de la IA en Recursos Humanos

El despliegue de sistemas de inteligencia artificial en el ámbito de gestión de personas plantea desafíos legales que van más allá del cumplimiento técnico de las normativas. La dimensión legal en este contexto debe entenderse como una herramienta para proteger derechos fundamentales y delimitar el poder de las tecnologías automatizadas, especialmente cuando estas afectan decisiones críticas sobre la vida profesional de las personas.

El marco jurídico existente ya ofrece algunos principios generales aplicables —como el derecho a la no discriminación, a la intimidad, a la información y al debido proceso—, pero la aplicación concreta de estos principios en entornos donde interviene la IA aún genera importantes zonas grises. Por ejemplo, ¿puede considerarse legalmente válida una decisión de no contratación basada en un sistema algorítmico entrenado con datos históricos de sesgo? ¿Qué grado de intervención humana se requiere para que una evaluación automatizada sea legítima? Estas preguntas han sido objeto de estudio en informes recientes sobre derechos fundamentales y automatización, como los publicados por la Agencia de los Derechos Fundamentales de la UE (FRA, 2020).

El Reglamento General de Protección de Datos introduce algunas salvaguardas, en particular respecto a la toma de decisiones automatizadas (Reglamento UE 2016/679). Sin embargo, estas disposiciones no abordan de forma específica el impacto de la IA en los procesos laborales, lo que ha dado lugar a interpretaciones divergentes y a un debate activo sobre la necesidad de regulación específica en el ámbito del empleo (ILO, 2021). Este vacío es parcialmente cubierto por propuestas como el RIA, que considera de “alto riesgo" muchos sistemas aplicados en el contexto de RR. HH (European Commission, 2021).

Desde una perspectiva organizacional, los aspectos legales de la IA en Recursos Humanos no solo exigen cumplimiento, sino también capacidad de interpretación y adaptación proactiva. Las empresas, además de anticipar las exigencias normativas actuales, también tienen que poner el foco en los posibles litigios futuros asociados a decisiones algorítmicas, el acceso desigual a oportunidades o la opacidad de ciertos sistemas.

Estudios recientes han advertido sobre la desconexión entre el desarrollo de tecnologías algorítmicas y la legislación antidiscriminación vigente (Wachter et al., 2021), lo que refuerza la necesidad de una gobernanza jurídica dinámica y multidisciplinar.

Por tanto, los aspectos legales no deben tratarse como una etapa final de validación, sino como un componente transversal del diseño, desarrollo y despliegue de la IA en RR. HH. Incluir la legalidad como eje desde las fases más tempranas del proceso favorece sistemas más robustos, auditables y socialmente aceptables, y contribuye a preservar la confianza tanto de los empleados como de la sociedad en su conjunto (De Stefano, 2022).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

b) Cumplimiento de las leyes laborales

La incorporación de la inteligencia artificial en los procesos de gestión de personas no solo implica retos tecnológicos o éticos, sino también la necesidad de alinearse con la normativa laboral vigente, que establece derechos y obligaciones fundamentales en el marco de la relación laboral. Aunque muchas leyes laborales fueron concebidas en un contexto analógico, sus principios centrales —como la igualdad de trato, la protección contra el despido injustificado, la intimidad y la libertad sindical— siguen siendo plenamente aplicables en entornos automatizados (Estatuto de los Trabajadores, 2015).

Uno de los principales desafíos es garantizar que el uso de sistemas de IA no derive en prácticas que vulneren estos derechos, aunque no exista intención explícita de hacerlo. Por ejemplo, un algoritmo de evaluación del rendimiento que clasifica automáticamente a los trabajadores según métricas de productividad podría incurrir en despidos o promociones sesgadas si no contempla los factores contextuales protegidos por la legislación laboral (De Stefano & Taes, 2021). Del mismo modo, los sistemas que rastrean comunicaciones internas o tiempos de actividad pueden colisionar con el derecho a la intimidad y con el principio de proporcionalidad en la vigilancia laboral (Ministerio de Trabajo, 2022).

Las organizaciones también deben considerar cómo se articula el uso de IA con el principio de no discriminación, tanto directa como indirecta, especialmente en los procesos de selección, promoción o asignación de tareas. Las leyes laborales prohíben expresamente cualquier forma de discriminación por razones de género, edad, origen étnico, orientación sexual, discapacidad o creencias, entre otras (Directiva 2000/78/CE). Si un sistema de IA reproduce patrones históricos sesgados sin mecanismos de corrección o supervisión, se corre el riesgo de incumplir con esta exigencia, aun cuando el algoritmo no utilice explícitamente variables protegidas (ETUC, 2022).

Otro punto crítico es el papel de la negociación colectiva y la representación de los trabajadores en la introducción de tecnologías que afectan sus condiciones laborales. La normativa reconoce el derecho de los trabajadores a ser informados y consultados cuando se introducen herramientas tecnológicas que puedan incidir en la organización del trabajo o en la seguridad y salud en el empleo (EFFAT, 2023). Ignorar estos procedimientos puede conllevar no solo conflictos laborales, sino también sanciones por parte de la autoridad competente.

Por tanto, cumplir con las leyes laborales en contextos de IA requiere algo más que validar técnicamente los algoritmos: exige una lectura jurídica sensible al contexto de uso y una colaboración activa entre los equipos de RRHH, los departamentos legales y los órganos de representación laboral. La legalidad no debe entenderse como un obstáculo para la innovación, sino como una condición de legitimidad, sostenibilidad y justicia organizacional (OIT, 2021).

c) Mitigación de riesgos legales

La implantación de modelos de IA en el ámbito de Recursos Humanos conlleva una serie de riesgos jurídicos específicos, que pueden derivarse tanto de fallos técnicos como de decisiones organizativas deficientes. La mitigación de estos riesgos no solo responde a una lógica de prevención de litigios, sino que constituye una condición clave para lograr una implementación legítima, equitativa y sostenible de la IA en el entorno laboral (FRA, 2020).

Entre los principales riesgos se encuentran la discriminación indirecta, la vulneración de la intimidad,

la desinformación de los trabajadores sobre el funcionamiento del sistema, la opacidad en los criterios de decisión y la ausencia de supervisión humana efectiva. Cada uno de estos elementos puede traducirse en una infracción de derechos fundamentales si no se actúa con diligencia. Por ejemplo, si un sistema de IA reproduce sesgos de género al predecir el potencial de liderazgo, o si se automatiza una decisión de despido sin ofrecer a la persona afectada la posibilidad de comprender o impugnar dicha decisión, se genera un riesgo legal de alta exposición (Veale & Edwards, 2018).

La identificación temprana de estos riesgos es esencial. Esto implica realizar una evaluación previa del impacto legal y ético de los sistemas de IA que se pretenden aplicar en RRHH, así como analizar las consecuencias que podrían tener sobre la carrera profesional, la privacidad y la autonomía de los empleados. Las evaluaciones de impacto en derechos fundamentales o en protección de datos —como las previstas en el RGPD o promovidas por autoridades como la AEPD o el ICO— ofrecen un marco útil (AEPD, 2020), (ICO, 2021), aunque no siempre suficiente si no se articulan con una cultura de revisión y aprendizaje continuo.

Otro aspecto relevante es la asimetría de información. En muchos casos, los trabajadores candidatos no son plenamente conscientes de la existencia de los sistemas algorítmicos ni de los datos que los alimentan, lo que impide el ejercicio efectivo de sus derechos. La falta de transparencia puede generar desconfianza y aumentar el riesgo de demandas por falta de consentimiento informado, sesgo no declarado o tratamiento indebido de datos (Jobin et al., 2019).

La interacción entre los sistemas de IA y la normativa laboral, de igualdad y de protección de datos hace necesaria una mirada jurídica integrada. Para mitigar riesgos legales, no basta con verificar la conformidad normativa de cada herramienta: se requiere una visión transversal que analice cómo interactúan las decisiones algorítmicas con las dinámicas organizativas, y que prevea mecanismos de corrección en caso de error, injusticia o desvío no intencionado (European Commission, 2021).

Por tanto, la mitigación de riesgos legales debe ser entendida como una práctica anticipatoria, participativa y multidisciplinar. Involucrar desde el inicio a juristas, responsables del talento, expertos en ética digital y representantes de los trabajadores permite diseñar sistemas más seguros, equitativos y jurídicamente sólidos, que refuercen la confianza en el uso de IA en el trabajo (Floridi et al., 2018).

d) Estrategias de cumplimiento normativo

A medida que los sistemas de IA se integran en los procesos de Recursos Humanos, el cumplimiento normativo deja de ser una cuestión reactiva para convertirse en un requisito estructural del diseño organizativo. Implementar estrategias eficaces de cumplimiento no solo evita sanciones legales, sino que fortalece la confianza, mejora la transparencia interna y contribuye a una cultura organizacional ética y responsable (AEPD, 2021).

Una primera estrategia clave es establecer protocolos de validación y revisión continua de los sistemas de IA utilizados. Estos protocolos deben permitir verificar que los algoritmos no generan resultados discriminatorios, que operan de forma trazable y que su funcionamiento puede ser explicado a las personas afectadas. El RIA establece requisitos concretos en esta dirección, como la documentación técnica, la supervisión humana y la transparencia del sistema (European Commission, 2021). Herramientas como el AI and data protection risk toolkit del ICO ofrecen marcos prácticos para aplicar estos controles en entornos reales (ICO, 2022).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Otra estrategia esencial es el desarrollo de políticas internas de uso responsable de IA, que definan claramente los límites, finalidades y responsabilidades asociadas a cada herramienta. Estas políticas deben estar alineadas con el marco legal vigente, pero también con los principios éticos establecidos por organizaciones como la OECD o el grupo de expertos de alto nivel de la Comisión Europea (AI HLEG), que recomiendan incorporar valores como la transparencia, la justicia y la supervisión activa desde la fase de diseño (OECD, 2021), (AI HLEG, 2019).

El rol de la formación también es central. La implementación de estrategias de cumplimiento exige que tanto los responsables de RRHH como los equipos técnicos y directivos comprendan los riesgos, obligaciones y buenas prácticas asociadas al uso de IA. Esto implica diseñar programas de capacitación adaptados a distintos niveles de responsabilidad, en línea con las recomendaciones de marcos como el de IEEE Ethically Aligned Design o el propuesto por COTEC para la gobernanza algorítmica en organizaciones (IEEE, 2020), (COTEC, 2022).

Además, debe asegurarse una canalización efectiva de dudas, quejas y reclamaciones relacionadas con los sistemas automatizados. Iniciativas como los 10 Principios para una IA ética de UNI Global Union subrayan la importancia de establecer estos canales internos para garantizar el respeto de los derechos laborales y digitales (UNI Global Union, 2021).

Por último, una estrategia avanzada de cumplimiento normativo debe integrar el enfoque de compliance multidisciplinar y evolutivo. Esto significa contar con comités o grupos de trabajo que reúnan perfiles jurídicos, técnicos, de gestión del talento y de representación laboral, capaces de supervisar y ajustar continuamente el uso de IA. Este modelo responde mejor a la complejidad regulatoria y permite adaptarse ágilmente a los cambios legislativos y a las exigencias sociales emergentes (Wachter et al., 2017).

2.5.7 Buenas prácticas y casos aplicados

a) Buenas prácticas para la gestión de riesgos

Las buenas prácticas en la gestión de riesgos derivados del uso de sistemas de IA en Recursos Humanos son fundamentales para convertir los principios ético-legales en procesos sostenibles, auditables y socialmente responsables. Estas prácticas, cuando se integran en la cultura organizativa, permiten anticipar problemas, mejorar la calidad de las decisiones y reforzar la confianza tanto interna como externa (COTEC, 2022).

Una primera buena práctica es la implementación de un enfoque proactivo de revisión algorítmica, que combine el uso de métricas de equidad, validación de variables sensibles y pruebas de robustez técnica con la supervisión humana constante. Organizaciones como IBM y LinkedIn han desarrollado marcos internos para el control del sesgo en herramientas de IA utilizadas en procesos de selección, evaluación o promoción, aplicando métricas comparativas entre grupos y procedimientos de validación cruzada (IBM, 2020), (LinkedIn, 2021).

Otra práctica es la implicación de actores en la Gobernanza de la IA. La conformación de comités o grupos internos multidisciplinares —con presencia de responsables de RR. HH., técnicos, juristas y representantes— ha demostrado ser eficaz para revisar decisiones críticas, interpretar resultados complejos y consensuar límites. Este enfoque ha sido adoptado por empresas como Telefónica, que ha creado consejos éticos internos para supervisar la IA con impacto en trabajadores (Telefónica, 2020).

También se considera buena práctica la transparencia hacia los empleados respecto a qué tecnologías se utilizan, qué datos se recogen, cómo se analizan y qué impacto pueden tener. Esto no solo cumple con principios legales como el consentimiento informado, sino que permite generar espacios de diálogo, comprensión y corresponsabilidad. La publicación de principios éticos, guías internas y habilitación de canales de consulta ha sido promovida tanto por actores institucionales como por organizaciones sindicales (UNI Global Union, 2020).

Por último, cabe destacar el valor de la evaluación de impacto previa al despliegue de sistemas de IA, que incluye aspectos técnicos, éticos, organizativos y legales. Estas evaluaciones, cuando son participativas y documentadas, permiten detectar posibles vulnerabilidades antes de que se conviertan en conflictos reales. Las guías de la Agencia Española de Protección de Datos y del ICO del Reino Unido han servido como base para diseñar estos protocolos en organizaciones públicas y privadas (AEPD, 2021), (ICO, 2022).

La sistematización de estas prácticas no solo mitiga riesgos, sino que facilita la escalabilidad del uso responsable de IA en RR. HH. Lejos de ser procedimientos burocráticos, estas acciones representan una inversión estratégica en sostenibilidad, reputación y calidad organizativa.

La utilidad de estas prácticas se refleja con claridad en distintos contextos reales, como se mostrará en el apartado siguiente sobre estudios de caso.

b) Análisis de casos sobre el uso ético y legal de la IA

El análisis de casos reales permite comprender cómo los principios éticos y las obligaciones legales se enfrentan a los desafíos concretos del uso de la IA en Recursos Humanos. Algunos ejemplos muestran esfuerzos responsables y proactivos, mientras que otros revelan fallos de gobernanza, opacidad o abuso algorítmico. Ambos tipos de casos ofrecen lecciones valiosas para empresas, legisladores y profesionales del ámbito laboral.

Uno de los casos más citados es el del sistema automatizado de reclutamiento de Amazon, que la empresa decidió dejar de utilizar tras detectarse que penalizaba sistemáticamente a las candidatas mujeres. El algoritmo había sido entrenado con datos históricos de contratación predominantemente masculina, lo que derivó en un sesgo de género indirecto (Dastin, 2018; Gray & Suri, 2019). Este caso puso en evidencia el riesgo de confiar acríticamente en sistemas que replican desigualdades estructurales, así como la necesidad de auditorías algorítmicas, trazabilidad y control humano en procesos sensibles como la selección de personal.

Casos similares de uso excesivo de tecnologías automatizadas han generado preocupaciones significativas en el ámbito de la vigilancia laboral algorítmica. En Estados Unidos, empleados de Amazon denunciaron los efectos del sistema Time Off Task (TOT), que mide el tiempo que un trabajador pasa fuera de su puesto de trabajo, señalando que esta vigilancia constante contribuye al agotamiento físico y mental y puede desembocar en sanciones o despidos automatizados (Nguyen, 2021; Gutelius & Pinto, 2023). En Alemania, algunos empleados de H&M cuestionaron la introducción de tecnologías como la identificación por radiofrecuencia (RFID) y sistemas de gestión algorítmica, alertando sobre una mayor precarización, intensificación del trabajo y pérdida de autonomía (Cole et al., 2022). Estas situaciones ilustran cómo la vulneración de la privacidad y la deshumanización de las decisiones pueden dañar la salud psicosocial y erosionar la confianza organizacional.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

En contraste, algunas organizaciones han mostrado un compromiso activo con el desarrollo ético de la IA. La compañía Accenture ha sido reconocida por establecer un marco interno de gobernanza algorítmica que incluye un comité transversal encargado de revisar la trazabilidad, explicabilidad y equidad de cada nuevo sistema aplicado a RRHH (Accenture, 2020; World Economic Forum, 2022). Además, la empresa ha priorizado la formación en IA ética entre sus empleados y ha habilitado canales de retroalimentación para evaluar el impacto de sus herramientas, configurando un ejemplo de implementación responsable y participativa.

También se han documentado reacciones institucionales correctivas en el sector público. En el Reino Unido, el Departamento de Trabajo y Pensiones diseñó un sistema de IA para detectar riesgos de fraude en prestaciones sociales. Tras denuncias por falta de transparencia y posibles discriminaciones, el sistema fue sometido a una revisión independiente y se establecieron mecanismos de publicación periódica sobre sus criterios, errores y tasas de acierto (UK Department for Work and Pensions, 2020; Ada Lovelace Institute, 2022). Este caso demuestra cómo, ante la presión social y legal, una administración puede reconducir una herramienta hacia estándares más legítimos, auditables y explicables.

Por último, el escándalo de la filtración de datos en el sitio web Ashley Madison, aunque no relacionado directamente con RR. HH., evidencia los riesgos críticos de una gestión inadecuada de datos personales sensibles. La exposición de información íntima de millones de usuarios muestra el impacto devastador que pueden tener las brechas de seguridad en la reputación organizacional y en la vida privada de las personas (Mansfield-Devine, 2015). Este caso refuerza la necesidad de políticas sólidas de protección de datos, especialmente en entornos de vigilancia corporativa y recopilación masiva de información biométrica, conductual o emocional.

Estos ejemplos, diversos en su origen y alcance, evidencian que el éxito o el fracaso en el uso de IA en RR. HH. no depende solo de la calidad técnica del sistema, sino del modelo de gobernanza adoptado, la participación de los actores implicados, y la capacidad de prevenir y corregir errores a tiempo. Incorporar estas lecciones puede ayudar a diseñar políticas más justas, sostenibles y centradas en la dignidad del trabajador.

Los casos mostrados refuerzan la importancia de aplicar con rigor los marcos de gobernanza y cumplimiento normativo desarrollados en los apartados anteriores.

3. INVESTIGACIÓN EMPÍRICA DE IA EN RECURSOS HUMANOS

La evolución de la inteligencia artificial en el ámbito de Recursos Humanos ha transformado significativamente la forma en la que se gestionan y optimizan los procesos relacionados con el talento humano. Desde la selección de personal hasta la evaluación del desempeño y la planificación de carrera, entre otros, los algoritmos y modelos predictivos han demostrado un enorme potencial para identificar patrones, anticipar necesidades y mejorar la eficiencia organizativa. Sin embargo, esta incursión de la IA en entornos laborales también plantea interrogantes sobre la equidad, la privacidad y el impacto en la toma de decisiones.

El presente capítulo aborda un enfoque empírico para el estudio de la inteligencia artificial en RR. HH., estructurado en dos grandes bloques: un análisis detallado de los procesos algorítmicos que sustentan estas tecnologías, y el estudio de casos reales en los que se han implementado sistemas inteligentes en distintas organizaciones. A través de una metodología rigurosa y una visión multidisciplinar, exploraremos cómo los datos son capturados, procesados y utilizados en la toma de decisiones, evaluando su impacto desde una perspectiva técnica, legal y ética.

Con esta aproximación, se pretende no solo comprender el estado actual de la IA en el área de gestión de personas., sino también identificar los desafíos y oportunidades que emergen en su aplicación práctica, proponiendo recomendaciones para una adopción responsable y equitativa de estas tecnologías.

3.1 Procesos de análisis de algoritmos

El análisis de algoritmos constituye el pilar fundamental para garantizar la efectividad y la transparencia en la aplicación de la inteligencia artificial en RR. HH. Dado el volumen y la sensibilidad de los datos manejados, resulta imprescindible evaluar la precisión de los modelos predictivos y auditar su comportamiento para detectar posibles sesgos, riesgos de discriminación y vulnerabilidades en la privacidad de los datos.

En esta sección, se examinarán los procesos clave para el análisis de algoritmos en RR. HH., abarcando desde la Gobernanza de datos y el enfoque empírico para la validación de modelos, hasta un estudio multidisciplinar que contemple aspectos técnicos, legales y éticos. Además, se profundizará en las metodologías de análisis cualitativo y cuantitativo para evaluar el uso de estas herramientas en entornos empresariales, así como en la necesidad de informes técnicos que documenten los hallazgos y faciliten la trazabilidad de las decisiones algorítmicas.

El objetivo es proporcionar un marco sólido y bien fundamentado para el estudio de los algoritmos aplicados en Recursos Humanos, destacando la importancia de un enfoque empírico y la adopción de buenas prácticas en su implementación.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

3.1.1 Gobernanza del dato

La arquitectura técnica que sostiene una IA auditable y responsable

En un entorno donde la inteligencia artificial media en la adopción de decisiones sensibles en Recursos Humanos —como la selección, promoción o desvinculación de personas— la Gobernanza del dato deja de ser una preocupación periférica y se convierte en un requisito estructural. No se trata únicamente de cumplir con obligaciones legales: hay que dotar a los sistemas de IA de una base técnica sólida que permita asegurar su trazabilidad, control y explicabilidad.

Desde una perspectiva técnica, la Gobernanza del dato puede definirse como el conjunto de principios, roles, procesos y herramientas que aseguran que los datos empleados en algoritmos sean de calidad, estén documentados, sean auditables y estén protegidos frente a usos no autorizados o sesgados. Esta definición trasciende el plano ético o normativo: sin Gobernanza del dato no puede haber IA confiable.

Normativas aplicables y marcos de referencia

El Reglamento europeo de Inteligencia Artificial establece con claridad en su artículo 10 la obligación de asegurar la calidad, trazabilidad y representatividad de los datos utilizados en sistemas de alto riesgo, como los que se aplican en la gestión de personas. Esta exigencia se alinea con los principios del GDPR, que impone restricciones sobre el uso de atributos sensibles y refuerza el principio de minimización del dato. A su vez, la nueva norma internacional ISO/IEC 42001, dedicada a sistemas de gestión para la IA, aporta un marco integral que vincula la Gobernanza del dato con la gestión del riesgo, la asignación de roles y la supervisión de terceros. Las mejores prácticas identificadas en organizaciones como E008 —referente en la implementación de modelos de gobernanza avanzados— demuestran que estas normativas son aplicables y, sobre todo, operativas cuando se despliegan mediante estructuras técnicas claras y una asignación explícita de responsabilidades.

Roles técnicos y arquitectura organizativa

Un gobierno del dato eficaz exige una arquitectura institucional bien definida. En los modelos más maduros, como es el caso de uno de los algoritmos analizados en una gran empresa energética en el marco del Proyecto IA+Igual[8], se ha identificado un despliegue claro de responsabilidades técnicas:

- **Data owners:** responsables funcionales del dato, que validan sus definiciones y aseguran su coherencia con los objetivos del área (por ejemplo, cómo se mide la rotación o el desempeño).
- **Data stewards:** encargados de la calidad, trazabilidad y limpieza del dato a lo largo de su ciclo de vida.
- **Comités transversales:** integrados por perfiles técnicos, legales, éticos y de ciberseguridad, que validan decisiones colegiadas sobre accesos y usos del dato.

[8] **Nota:** Se ha elaborado un Informe Técnico confidencial que recoge 7 ejemplos de análisis algoritmo sobre casos de uso reales de modelos de IA en RR. HH -E001-E007- y 3 ejemplos centrados en la Gobernanza del dato en el mismo ámbito -E008-E010-. Los investigadores han utilizado un modelo de análisis metodológico propio, focalizado en la búsqueda de la equidad, para cumplir el objetivo del Proyecto IA+Igual de prevenir posibles sesgos. Este capítulo se basa en el documento anonimizado: el siguiente subapartado expone un resumen esquemático.

Esta combinación de roles permite evitar decisiones aisladas, garantizar la coherencia semántica y estructurar una cadena de custodia clara en entornos automatizados.

Control técnico del ciclo de vida del dato

La gobernanza técnica del dato se despliega a lo largo de todo su ciclo de vida. Desde la captura en origen, donde errores sistemáticos pueden introducir sesgos, hasta las fases de transformación, modelización y almacenamiento, se requiere control exhaustivo.

Las organizaciones más avanzadas trabajan con:

- Diccionarios de datos compartidos para asegurar una semántica común.
- Versionado de datos y registros de transformación.
- Sistemas de auditoría interna que validan la consistencia y representatividad de las variables utilizadas.

Mecanismos de trazabilidad, explicabilidad y supervisión

Uno de los pilares de la gobernanza técnica es la capacidad de explicar y supervisar el comportamiento de los algoritmos. Esto requiere, en primer lugar, que los datos empleados sean trazables. En segundo lugar, implica contar con herramientas de control: dashboards, indicadores de equidad y revisión algorítmica por terceros.

Estas son algunas herramientas:

- **Inventarios de soluciones IA.**
- **Formularios técnicos** de validación de proveedores.
- **Dashboards** de impacto algorítmico.
- **Fairness Indicators** con seguimiento por subgrupos.

Estos mecanismos de control se ven reforzados por soluciones como Warden AI, desarrollados para realizar auditorías continuas de sesgos con segmentación demográfica y criterios de alto riesgo, especialmente útiles en procesos de RR. HH. Estas herramientas permiten no solo vigilar errores, sino también analizar el comportamiento diferenciado de un sistema entre colectivos con distinta representación o características.

Además, organizaciones como HireVue han implementado declaraciones de explicabilidad pública con criterios diferenciados por audiencias —como comités sindicales, técnicos o usuarios—, mostrando que es posible establecer estándares operativos para explicar decisiones algorítmicas en contextos laborales sensibles.

El uso de sandboxes éticos para testear el impacto de los algoritmos en condiciones controladas ha demostrado ser una práctica eficaz para anticipar consecuencias no deseadas sin comprometer datos reales.

Conclusión: una capa técnica para una IA justa

El desarrollo de IA en RR. HH. no puede desligarse de una gobernanza técnica robusta. Lejos de ser

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

un complemento burocrático, es el sistema nervioso que convierte un modelo estadístico en una herramienta responsable. Permite auditar decisiones, corregir sesgos y garantizar el respeto a los derechos de las personas evaluadas.

Firmas como IBM han articulado marcos de gobernanza técnica que integran, por ejemplo, el diseño de la trazabilidad, la asignación de roles y los criterios de control de ciclo de vida, ofreciendo modelos replicables y alineados con normas internacionales, según se refleja en el informe sobre gobernanza elaborado a partir del caso E008.

A su vez, estudios como los de AIHR Institute y HR Dive han evidenciado la necesidad de monitorizar no solo errores cuantificables, sino también el impacto percibido por parte de trabajadores y responsables sobre la equidad y fiabilidad del sistema.

Sin datos trazables, sin supervisión colegiada y sin control técnico del ciclo de vida, la IA no solo es opaca: es inviable desde el punto de vista normativo y ético. La gobernanza técnica es, por tanto, la base silenciosa pero indispensable para una IA explicable, auditada y aceptada socialmente.

3.1.2 Un enfoque empírico para analizar la IA en RR. HH.

Del modelo predictivo a la evidencia verificable

A medida que los sistemas de inteligencia artificial se integran en los procesos de gestión de personas, crece la necesidad de someter estas soluciones a un escrutinio riguroso. Un enfoque empírico aplicado a la IA en Recursos Humanos implica no limitarse a desarrollar modelos teóricamente sólidos, sino demostrar con evidencia medible cómo se comportan estos sistemas en entornos reales y qué consecuencias producen sobre los individuos y las organizaciones (Wilson, Hoffman & Morgenstern, 2021).

Este enfoque se basa en tres pilares fundamentales: la validación cuantitativa de los modelos, la interpretación comprensible de sus decisiones y la capacidad de auditar su funcionamiento de forma sistemática (Varshney, 2022). Así, la investigación empírica se convierte en un mecanismo de verificación, ajuste y confianza, especialmente relevante cuando los algoritmos inciden en trayectorias profesionales, oportunidades laborales o evaluaciones de desempeño.

Herramientas y metodologías para la evaluación empírica

En el plano cuantitativo, las técnicas de validación cruzada como k-fold, leave-one-out o hold-out split son esenciales para estimar la generalización del sistema a nuevos datos (Wilson et al., 2021). En entornos de RR. HH., permiten verificar si un modelo mantiene su precisión cuando se aplican cohortes distintas, cambios temporales o variaciones contextuales.

A este tipo de validación se suman herramientas de interpretabilidad que desarrollaremos en el apartado “Interpretabilidad algorítmica: comprender cómo decide un modelo” del punto 3.1.5., que permiten analizar cómo influye cada variable en la decisión del modelo, incluso en sistemas opacos como los de tipo black-box. Estas técnicas contribuyen a generar explicaciones comprensibles tanto para usuarios técnicos como para perfiles no especializados, un aspecto que ha sido identificado como clave desde la perspectiva regulatoria (Binns, 2018).

Además, han emergido marcos técnicos para la evaluación de sesgos y equidad, como AI Fairness 360 (IBM, 2023) o Warden AI (2024), que permiten realizar auditorías algorítmicas mediante el análisis de métricas diferenciadas por subgrupos poblacionales y el monitoreo de tasas de error por colectivo.

Ejemplo ilustrativo: auditoría de un algoritmo de selección de personal

Imaginemos un sistema automatizado que clasifica currículums para decidir qué candidaturas avanzan a una entrevista. Una auditoría empírica de este modelo podría comenzar con la evaluación de su rendimiento técnico: se aplicaría una validación cruzada para comprobar la estabilidad del sistema y se analizarían métricas como la precisión (porcentaje de aciertos) o la curva AUC-ROC, que mide su capacidad para distinguir entre perfiles adecuados e inadecuados (Wilson et al., 2021).

En una segunda fase, se identificarían las variables que más influyen en la toma de decisiones del modelo. Este paso permite verificar si el sistema está priorizando factores relevantes —como la experiencia o la formación— o si, por el contrario, se están filtrando elementos no justificados (como proxies indirectos de edad o género).

A continuación, se evaluaría la posible presencia de sesgos. Para ello, podrían emplearse herramientas como AI Fairness 360 (aif360), que permiten comparar cómo se comporta el algoritmo en distintos subgrupos. Por ejemplo, si detectamos que las mujeres con perfiles equivalentes reciben sistemáticamente menos oportunidades que los hombres, se trataría de un sesgo en la tasa de falsos negativos (IBM, 2023).

Paralelamente, se integrarían mecanismos de interpretabilidad que ayuden a comprender cómo toma decisiones el modelo. Algunas plataformas comerciales, como HireVue o Microsoft Responsible AI, han desarrollado entornos que adaptan el nivel de explicación al tipo de usuario: un técnico de selección puede recibir detalles sobre la lógica del modelo, mientras que una persona candidata podría obtener una versión simplificada y comprensible del porqué fue descartada.

Por último, se comprobaría si el sistema permite intervención humana: es decir, si un responsable del proceso puede revisar, corregir o revertir decisiones automatizadas en casos conflictivos. Esta capacidad de reversibilidad es esencial para mantener la agencia humana y evitar decisiones arbitrarias (Doshi-Velez & Kim, 2017).

Metodologías complementarias para validar decisiones automatizadas

Además de la auditoría técnica, algunos enfoques incorporan validación contextual y ética mediante paneles interdisciplinarios que revisan muestras de decisiones automatizadas para garantizar su coherencia normativa, técnica y organizativa (World Economic Forum, 2022).

Esta validación se complementa con pruebas de estrés algorítmico, diseñadas para simular escenarios de entrada atípicos y analizar la estabilidad del sistema (Almeida et al., 2023).

Modelos más avanzados han incorporado también interpretabilidad conceptual, como las Concept Activation Vectors (Google Research, 2023), que permiten verificar si el modelo algorítmico está respondiendo a constructos relevantes desde el punto de vista del dominio profesional.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Conclusión: una vía hacia la confianza técnica y organizativa

Adoptar un enfoque empírico no es solo una exigencia metodológica, sino una condición para construir sistemas de IA legítimos y confiables en el ámbito de los Recursos Humanos (European Comission, 2021). A través de la validación empírica, las organizaciones pueden reducir incertidumbres, anticipar impactos negativos y demostrar el cumplimiento de principios fundamentales como la equidad, la explicabilidad y el control humano significativo.

Este tipo de análisis requiere, como se verá en el siguiente apartado, una perspectiva multidisciplinar que articule la dimensión técnica con marcos legales, éticos y organizativos (Wachter, Mittelstadt & Floridi, 2017).

3.1.3 La necesidad de una visión multidisciplinar

Comprender la IA en RR. HH. más allá de lo técnico

La aplicación de inteligencia artificial en Recursos Humanos exige algo más que sofisticación algorítmica. No basta con desarrollar modelos estadísticamente eficientes o sistemas predictivos de última generación. La naturaleza del entorno laboral, profundamente vinculada a derechos fundamentales, expectativas humanas y contextos organizativos complejos, impone una lectura multidisciplinar de cualquier herramienta basada en IA (Bodie et al., 2022; De Stefano & Aloisi, 2018).

Así, los modelos de IA utilizados no pueden abordarse únicamente desde la ingeniería o la ciencia de datos. Debe analizarse desde la intersección de al menos cuatro dimensiones: técnica, legal, ética y especializada en gestión de personas. Cada una aporta una perspectiva necesaria, y su integración es lo que permite desarrollar tecnologías legítimas, sostenibles y alineadas con los valores democráticos que rigen el mundo del trabajo (Floridi & Cowls, 2022; ILO, 2022).

a) Propuesta de marco multidisciplinar

Dimensión técnica: modelos, algoritmos y NLP aplicados al talento

Desde el plano técnico, la IA en RR. HH. se estructura a partir de modelos predictivos, algoritmos de clasificación y técnicas de procesamiento del lenguaje natural. Estas herramientas permiten automatizar tareas como el cribado de candidaturas o la predicción de rotación laboral (Akool, 2024). Para hacer comprensibles las decisiones algorítmicas, se recurre a métodos de interpretabilidad que ayudan a identificar qué variables están influyendo en los resultados, facilitando explicaciones comprensibles también para profesionales sin perfil técnico. (Doshi-Velez & Kim, 2017).

El reto consiste en mantener la eficiencia algorítmica sin comprometer la transparencia operativa ni la auditabilidad del sistema (Morley et al., 2021).

Dimensión legal: derechos, límites y equilibrio de intereses

Desde la perspectiva jurídica, la IA en el trabajo debe ajustarse a marcos como el Reglamento General de Protección de Datos y el RIA. Uno de los puntos más sensibles es el uso del interés legítimo como base legal para procesar datos en procesos automatizados. Si bien está contemplado en el artículo 6.1.f

del GDPR, su aplicación en contextos laborales exige un equilibrio estricto entre los objetivos organizativos y los derechos individuales, bajo principios de proporcionalidad y necesidad (Wachter, Mittelstadt & Floridi, 2017).

Además, aspectos como el derecho a la explicación o la intervención humana en decisiones automatizadas siguen siendo objeto de debate jurídico (Binns, 2018), especialmente en relación con el impacto de estas herramientas en la autonomía del trabajador.

Dimensión ética: justicia, sesgos y explicabilidad

La ética de la inteligencia artificial ha evolucionado hacia un marco común de principios, entre ellos el principio de no causar daño, la justicia, la autonomía y la explicabilidad (Floridi & Cowls, 2022; Jobin, Ienca & Vayena, 2019). En RR. HH., la equidad en la toma de decisiones es una prioridad crítica, especialmente en sistemas entrenados con datos históricos que pueden reflejar o amplificar sesgos existentes (Mittelstadt et al., 2016).

La ética aplicada demanda que las organizaciones no solo implementen sistemas técnicamente correctos, sino también moralmente justificables, incluyendo procesos de revisión participativos y estructuras que permitan cuestionar decisiones algorítmicas (ILO, 2022).

Dimensión especializada: saberes y prácticas de RR. HH.

La incorporación de IA en RR. HH. no puede hacerse sin la mediación del conocimiento experto en gestión del talento. Organizaciones como el World Economic Forum (2021) han subrayado la necesidad de enfoques centrados en las personas, donde la automatización respete los procesos organizativos y los valores culturales de cada empresa. Las decisiones de contratación, promoción o desvinculación deben estar contextualizadas, considerando factores que escapan al modelo matemático, como la experiencia informal o la compatibilidad con equipos.

RR. HH. también actúa como puente entre el diseño técnico y el impacto organizativo, garantizando que los modelos sean útiles, comprensibles y aceptables para los colectivos afectados (SHRM, 2023).

Una mirada comparativa: riesgos y medidas desde cada dimensión

Tabla 18: *Riesgos y medidas correctivas por cada una de las cuatro dimensiones clave de IA*

Dimensión	Riesgo detectado	Medida correctiva propuesta
Técnica	Modelo sobre ajustado o inestable	Validación cruzada, revisión continua del rendimiento
Legal	Tratamiento desproporcionado de datos personales	Test de proporcionalidad, intervención humana significativa
Ética	Reproducción de sesgos históricos	Evaluación de equidad algorítmica, explicabilidad
RR. HH.	Resultados no alineados con la cultura y estrategia	Supervisión funcional, contextualización organizativa

Fuente: elaboración propia.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Conclusión: hacia una integración madura de saberes

Integrar la inteligencia artificial en los procesos de gestión de personas sin considerar sus implicaciones legales, éticas y organizativas es un error que muchas empresas están empezando a comprender. Una visión multidisciplinar no es solo deseable, sino necesaria para garantizar la legitimidad, efectividad y aceptabilidad de estos sistemas (World Economic Forum, 2021).

Cuando estas dimensiones trabajan de forma coordinada, es posible construir soluciones de IA que no solo sean técnicamente competentes, sino también equitativas, transparentes y humanas.

3.1.4 Análisis cualitativo

La voz de las personas como evidencia en la evaluación de la IA

En la evaluación de sistemas de inteligencia artificial aplicados a Recursos Humanos, los datos cuantitativos —como métricas de precisión, tasas de error o medidas de rendimiento— no son suficientes para comprender el impacto real de estas tecnologías. La forma en que los trabajadores, responsables y candidatos perciben y experimentan el uso de IA en procesos como la selección, la evaluación o la promoción también ofrece información crucial. Este tipo de conocimiento no puede captarse con modelos estadísticos: requiere métodos de análisis cualitativo (Tracy, 2010).

El análisis cualitativo permite acceder a las dimensiones subjetivas, organizacionales y contextuales que rodean la implementación de tecnologías inteligentes en el ámbito laboral. A través de entrevistas, focus groups o encuestas abiertas, se pueden identificar inquietudes, resistencias, percepciones de justicia o legitimidad, así como prácticas informales que escapan a los registros estructurados (Guest, Namey & Mitchell, 2013). Incorporar estas voces en el proceso de evaluación mejora la legitimidad social de la IA, permite anticipar barreras culturales y ayuda a diseñar intervenciones más ajustadas a la realidad de los entornos de trabajo (Van den Broek et al., 2021).

Herramientas y metodologías para un análisis en profundidad

El análisis cualitativo en inteligencia artificial aplicada a Recursos Humanos y al ámbito laboral se basa en métodos inductivos, interpretativos y experienciales (Flick, 2018). Entre las estrategias más empleadas se encuentran:

- **Entrevistas semiestructuradas,** que permiten profundizar en la experiencia personal de los usuarios, tanto en la fase de adopción como en el uso continuado del sistema.
- **Focus groups,** donde se exploran de forma colectiva las percepciones compartidas sobre el impacto de la IA en dinámicas como la transparencia, la autonomía o la percepción de control.
- **Encuestas abiertas,** útiles para captar comentarios espontáneos y expresiones emocionales, que luego pueden codificarse temáticamente (Braun & Clarke, 2006).

Estas técnicas permiten recoger datos ricos en matices, que pueden analizarse mediante software especializado como NVivo (Bazeley & Jackson, 2013) o ATLAS.ti (Silver & Lewins, 2014), facilitando procesos de codificación, segmentación temática y visualización de patrones. Estas herramientas permiten crear categorías emergentes, mapear conceptos clave o incluso generar modelos de percepción que complementan los resultados cuantitativos (Kuckartz, 2019).

En investigaciones aplicadas, este enfoque ha demostrado ser especialmente útil en procesos de selección automatizada, donde los candidatos pueden sentirse deshumanizados, o en evaluaciones de desempeño apoyadas por IA, en las que los empleados cuestionan la legitimidad del criterio algorítmico (Oostrom, Kogan & Born, 2020; Robertson, 2021). En estos contextos, los estudios de caso han revelado cómo las tecnologías no sólo transforman procesos, sino también culturas organizacionales, jerarquías implícitas y dinámicas de poder (Park & Humphry, 2019).

El papel de la percepción: legitimidad, confianza y adaptación

Los datos cualitativos también permiten identificar percepciones de equidad, comprensión o exclusión por parte de los distintos actores involucrados. Por ejemplo, mánagers que no comprenden los criterios utilizados por el algoritmo pueden resistirse a su implementación, mientras que empleados que no reciben una explicación adecuada quizá desconfíen del sistema incluso si sus decisiones son técnicamente correctas (Van den Broek et al., 2021; Robertson, 2021).

Estas percepciones afectan directamente la adopción y sostenibilidad de los sistemas de IA en entornos laborales y, por tanto, deben ser tenidas en cuenta como insumo diagnóstico y guía para rediseñar la herramienta, ajustar la comunicación o fortalecer el acompañamiento humano en los procesos automatizados (Tracy, 2010).

Conclusión: del dato subjetivo al diálogo con lo cuantitativo

Los análisis cualitativos no deben verse como opuestos al enfoque cuantitativo, sino como su complemento necesario. Mientras que los datos numéricos permiten evaluar el rendimiento de un algoritmo, los relatos, opiniones y emociones expresadas por los usuarios permiten entender cómo estos sistemas son vividos e interpretados en la práctica (Charmaz, 2014; Miles, Huberman & Saldaña, 2014).

En los siguientes apartados, dedicados al análisis algorítmico, se verá cómo esta combinación entre lo cuantitativo y lo cualitativo permite obtener una evaluación más completa, anticipar resistencias, diseñar protocolos más robustos y legitimar las decisiones en el ámbito de la gestión de personas asistida por IA.

3.1.5 Análisis de algoritmos

a) Introducción: evaluar el corazón técnico de la IA en RR. HH.

En un contexto donde los sistemas de inteligencia artificial influyen ya en decisiones organizativas clave —como la contratación, la promoción o la desvinculación de personas—, el análisis técnico de los algoritmos deja de ser un ejercicio reservado a ingenieros para convertirse en un requisito transversal de gobernanza y responsabilidad institucional (AI Now Institute, 2019).

Los algoritmos aplicados a Recursos Humanos no son entes abstractos: están entrenados sobre datos históricos, optimizados según determinadas funciones objetivo y aplicados en entornos sociales marcados por desigualdades previas. Por tanto, evaluar su desempeño técnico no se limita a comprobar si "funcionan": requiere, además, analizar cómo y para quién funcionan (Barocas, Hardt & Narayanan, 2019).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Esto exige examinar métricas de rendimiento (como precisión o recall) e indicadores de equidad, explicabilidad, trazabilidad y robustez. En otras palabras, el análisis algorítmico en RR. HH. debe ir más allá de lo cuantitativo: ha de integrar el impacto que tiene sobre la justicia organizativa, la igualdad de oportunidades y la legitimidad institucional (Doshi-Velez & Kim, 2017).

Uno de los principales retos que plantea esta tarea es la posibilidad de que los algoritmos —al ser entrenados con datos históricos— reproduzcan o incluso amplifiquen desigualdades ya existentes, creando un ciclo vicioso de sesgo difícil de romper (Mehrabi et al., 2021). De ahí que el análisis técnico de los modelos sea también un análisis ético y estratégico (Raji & Buolamwini, 2019).

Este subapartado presenta las herramientas y criterios necesarios para llevar a cabo ese análisis con profundidad, responsabilidad y visión multidisciplinar. Se parte de los fundamentos del sesgo algorítmico y sus tipologías, se explican las principales métricas de equidad y evaluación, se describen las estrategias de mitigación más relevantes y se culmina con un ejemplo práctico aplicado a un sistema de selección automatizado.

b) Fundamentos del sesgo algorítmico

Comprender los fundamentos del sesgo algorítmico

Hablar de sesgo algorítmico no significa que los modelos "opinen" o "discriminen" como lo haría una persona. Significa que, como herramientas estadísticas entrenadas sobre datos del mundo real, los algoritmos pueden reproducir, amplificar o encubrir desigualdades existentes, sin ser conscientes de ello ni poder corregirse por sí mismos (Barocas, Hardt & Narayanan, 2019).

Un sesgo algorítmico puede definirse como una distorsión sistemática e injustificada en el comportamiento de un sistema automatizado que afecta de forma diferencial a ciertos grupos o individuos. No es lo mismo que un error aleatorio: se trata de un patrón persistente, predecible y estructural (Mehrabi et al., 2021).

Este tipo de sesgos puede tener consecuencias graves, especialmente en contextos sensibles como el ámbito laboral. La IA en RR. HH. puede tomar ayudar a decisiones que influyen directamente en las trayectorias profesionales de las personas: acceso al empleo, promociones, despidos, evaluaciones o salarios. Si esas decisiones están sesgadas, los efectos no solo son injustos, sino que pueden perpetuar o institucionalizar discriminaciones históricas (Eubanks, 2018).

¿Dónde y cómo surge el sesgo?

El sesgo no aparece únicamente "en los datos" ni "en el algoritmo". Es un fenómeno multicausal que puede surgir en cualquiera de las siguientes fases (Sandvig et al., 2014):

- **Sesgo en los datos de entrada:** Por ejemplo, si el conjunto de entrenamiento está desbalanceado o refleja decisiones humanas previas que ya eran discriminatorias.
- **Sesgo en la construcción del modelo:** Ocurre cuando el diseño del sistema —elección de variables, funciones de coste, umbrales— favorece ciertos perfiles o penaliza a otros sin justificación técnica ni ética.
- **Sesgo en la implementación y uso:** Puede darse cuando el modelo se aplica en contextos para los que no fue validado, o cuando se combina con procesos organizativos que lo sesgan aún más.

Tipologías funcionales del sesgo algorítmico

Existen múltiples formas en las que el sesgo puede manifestarse. Algunas de las más relevantes en entornos de RR. HH. son (véase tabla completa de sesgos algorítmicos):

• **Sesgo de representación**: Algunos grupos están infrarrepresentados en los datos, lo que reduce la precisión del modelo para esos perfiles.

• **Sesgo de selección:** El modelo aprende patrones previos de selección que pueden haber sido discriminatorios.

• **Sesgo de exclusión:** Se omiten variables clave o se usan proxies que ocultan dimensiones sociales relevantes.

• **Sesgo de medición:** La variable objetivo (por ejemplo, "rendimiento" o "éxito") está mal definida o refleja valores cuestionables. Un ejemplo paradigmático de este tipo de sesgo fue documentado por Obermeyer et al. (2019) en el ámbito sanitario estadounidense. El algoritmo analizado debía identificar a pacientes que necesitaban atención médica intensiva, pero usaba como variable objetivo el coste sanitario previo, asumiendo que mayor gasto indicaba mayor necesidad. Sin embargo, debido a desigualdades estructurales en el acceso a la salud, las personas negras recibían menos atención médica, por lo que su "coste" aparecía artificialmente más bajo. Como resultado, el modelo subestimaba sistemáticamente sus necesidades clínicas, reproduciendo una discriminación racial encubierta bajo un criterio económico erróneo. Este caso demostró cómo una mala elección de variable puede ocultar sesgos estructurales bajo una aparente neutralidad técnica.

• **Sesgo de evaluación:** Se aplican métricas globales sin desagregarlas por subgrupos, ocultando desigualdades estructurales.

Estas tipologías no son excluyentes. De hecho, suelen combinarse y retroalimentarse, generando distorsiones difíciles de detectar si no se analizan explícitamente (Binns, 2018).

El ciclo vicioso del sesgo

Uno de los mayores peligros de los sistemas algorítmicos es que pueden aprender sesgos del pasado y proyectarlos hacia el futuro. Si no se interviene, esto genera un ciclo vicioso (Barocas et al., 2019):

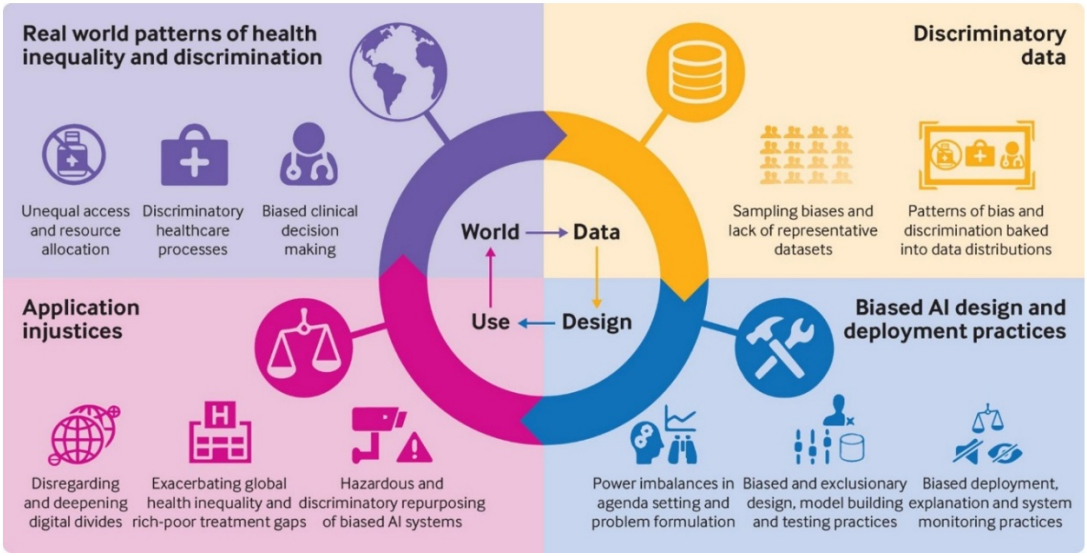
1. **Datos históricos** con desigualdades.
2. **Modelo entrenado** con esos datos.
3. **Resultados sesgados** aplicados a nuevas personas.
4. **Retroalimentación** que refuerza los patrones previos.

Aunque la grafía de Leslie et al. que se muestra a continuación se centra en el impacto de la inteligencia artificial y los modelos algorítmicos en la Salud tras el COVID 19, sirve de referencia.

Este ciclo convierte el sesgo en un elemento estructural del sistema, naturalizando resultados injustos bajo una apariencia de objetividad técnica.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Gráfico 2: *¿Puede incrementar la IA la desigualdad?*



Fuente: BMJ 2021; vol. 372, n 304.

c) Clasificación formal de la equidad y fairness metrics

Formalizar la equidad: criterios y métricas para evaluar justicia algorítmica

Una vez comprendido qué es el sesgo y cómo puede surgir, el siguiente paso es determinar cómo evaluarlo de forma empírica y cuantificable. Para ello, el campo de la inteligencia artificial ha desarrollado un conjunto de métricas llamadas fairness metrics, que buscan traducir distintas nociones de equidad en términos matemáticos (Verma & Rubin, 2018).

Estas métricas permiten diagnosticar cuándo un sistema automatizado está generando resultados injustos o desiguales, y en qué medida.

Sin embargo, medir la equidad no es una tarea neutral ni sencilla. No existe una única definición de justicia, ni todas las aproximaciones son compatibles entre sí. Según la filosofía política y la teoría de la probabilidad, las métricas de equidad pueden agruparse en tres familias conceptuales: independencia, separación y suficiencia (Barocas, Hardt & Narayanan, 2019).

Clasificación de métricas de equidad por familia conceptual

1. Independencia

Este criterio parte de la idea de que la decisión del algoritmo debe ser independiente de atributos sensibles como género, edad o etnia.

El enfoque está vinculado a ideas de igualdad de trato formal y de no discriminación directa (Binns, 2018). Aunque fácil de calcular, suele ser cuestionado cuando existen diferencias reales en los perfiles de desempeño entre grupos (Friedler et al., 2019).

Métrica	Fórmula
Demographic Parity	$P(\hat{Y}=1 \mid A=a) = P(\hat{Y}=1 \mid A=b)$
Statistical Parity Difference	$\text{SPD} = P(\hat{Y}=1 \mid A=a) - P(\hat{Y}=1 \mid A=b)$
Impact Ratio (4/5 rule)	$\text{Impact Ratio} = \frac{P(\hat{Y}=1 \mid A=a)}{P(\hat{Y}=1 \mid A=b)}$

Su métrica más conocida es la paridad demográfica (demographic parity), que exige que todos los grupos reciban decisiones positivas en la misma proporción.

2. Separación

La separación propone que el algoritmo puede tomar decisiones distintas para distintos grupos, pero no debe cometer más errores con unos que con otros. Por eso, estas métricas requieren que la predicción esté condicionada al resultado real. Lo importante es que el modelo cometa errores de forma similar entre grupos.

Está alineada con la idea de igualdad de acceso a oportunidades (Dwork et al., 2012). Su implementación suele preferirse en contextos donde se busca minimizar desigualdades de trato en quienes sí cumplen los criterios objetivos de éxito (Chouldechova, 2017).

Métrica	Fórmula
Equalized Odds	Para que se cumpla entre dos grupos: A = a y A = b, deben cumplirse las dos condiciones
True Positive Rate Parity Métrica de rendimiento diferencial	$P(\hat{Y}=1 \mid Y=1, A=a) = P(\hat{Y}=1 \mid Y=1, A=b)$
False Positive Rate Parity Predictive Equality	$P(\hat{Y}=1 \mid Y=0, A=a) = P(\hat{Y}=1 \mid Y=0, A=b)$
Equal Opportunity Criterio de equidad	$P(\hat{Y}=1 \mid Y=1, A=a) = P(\hat{Y}=1 \mid Y=1, A=b)$

Su métrica más conocida es Equalized odds y es útil cuando se busca justicia en el error.

3. Suficiencia

Este criterio exige que la decisión del modelo sea igualmente confiable para todos los grupos. Es decir, si un modelo predice que una persona será apta, esa predicción debería tener igual validez sin importar su grupo.

Este enfoque refleja una visión meritocrática basada en resultados, aunque puede entrar en tensión con Separation si las prevalencias difieren entre subgrupos (Kleinberg, Mullainathan & Raghavan, 2016).

La suficiencia exige confianza simétrica en el modelo, independientemente de la distribución real entre grupos.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Métrica	Fórmula
Predictive Parity	$P(Y=1 \mid \hat{Y}=1, A=a) = P(Y=1 \mid \hat{Y}=1, A=b)$
False Omission Rate Parity (FOR Parity)	$P(Y=1 \mid \hat{Y}=0, A=a) = P(Y=1 \mid \hat{Y}=0, A=b)$
Negative Predictive Value Parity	$P(Y=0 \mid \hat{Y}=0, A=a) = P(Y=0 \mid \hat{Y}=0, A=b)$
False Discovery Rate Parity	$P(Y=0 \mid \hat{Y}=1, A=a) = P(Y=0 \mid \hat{Y}=1, A=b)$
Calibration within Groups	$P(Y=1 \mid \hat{Y}=\hat{y}, A=a) = \hat{y}, \quad \forall a$

Su métrica más conocida es la Predictive parity, cuando el foco es la confianza en los aciertos del modelo.

Guía sobre notación:

Y Etiqueta verdadera.	$\hat{Y}$ Predicción del modelo.	A Atributo sensible.
Y = 1 **Verdadero positivo**	$\hat{Y} = 1$ resultado positivo	a, b categorías distintas
Y = 0 **Verdadero negativo**	$\hat{Y} = 0$ resultado negativo	

Tensiones entre métricas: ¿Se puede ser justo con todos los criterios?

Una de las lecciones más importantes de este bloque es que no es posible cumplir simultáneamente con los tres criterios cuando existen diferencias reales entre grupos. Esta "tríada imposible" ha sido demostrada formalmente en trabajos seminales (Kleinberg et al., 2016; Chouldechova, 2017), y pone de manifiesto que elegir una métrica es también una decisión normativa.

Es importante saber que Separación y Suficiencia son mutuamente excluyentes cuando las tasas base difieren entre grupos. Esto quiere decir que no se puede cumplir simultáneamente con ambas familias de métricas en muchos contextos, lo que obliga a tener que elegir según prioridades éticas, legales o prácticas.

Por ejemplo, mientras un sistema de justicia puede optar por separation para asegurar trato equitativo en los aciertos, una empresa puede inclinarse por suficiencia para asegurar coherencia en sus predicciones. En todos los casos, debe existir transparencia sobre el criterio adoptado (Berk et al., 2018).

Este marco conceptual es clave para interpretar correctamente las métricas cuantitativas de fairness, y será la base sobre la que se construya la comparación entre tipos de sesgo, métricas concretas y mitigaciones, que veremos en el siguiente bloque.

d) Tabla de equivalencia: tipo de sesgo – métrica – mitigación

De la teoría a la intervención: traducir sesgos en métricas y estrategias

Comprender los distintos tipos de sesgo y los enfoques formales de equidad es indispensable, pero no suficiente para intervenir con eficacia. En la práctica, los equipos técnicos y responsables de auditoría algorítmica necesitan traducir esas nociones en acciones concretas.

Este bloque presenta una tabla de equivalencias operativas que relaciona:

- **El tipo de sesgo identificado,** basado en las tipologías funcionales descritas previamente.
- **La métrica de fairness más adecuada para detectarlo o cuantificarlo** (Verma & Rubin, 2018; Barocas, Hardt & Narayanan, 2019).
- **La técnica o enfoque de mitigación que permite corregirlo o reducir su impacto,** considerando estudios comparativos como los de Friedler et al. (2019).

Esta tabla no pretende ser exhaustiva ni prescriptiva. La relación entre sesgos, métricas y mitigadores no es unívoca, y depende del contexto, los datos y los valores organizativos. Aun así, proporciona un mapa útil para orientar auditorías, evaluar decisiones técnicas y documentar procesos de intervención (Holstein et al., 2019).

Tabla 19: *Equivalencia según el tipo de sesgo, métrica de detección y técnica de mitigación*

Tipo de sesgo	Métrica de detección	Técnica o enfoque de mitigación
Representación	Paridad demográfica Skew ratio (Verma & Rubin, 2018)	Reponderación de clases, oversampling de grupos infrarrepresentados (Kamiran & Calders, 2012)
Selección	Equal opportunity odds (Hardt, Price & Srebro, 2016)	Ajuste del clasificador, regularización de errores condicionales (Friedler et al., 2019)
Exclusión	Fairness Through Awareness (Dwork et al., 2012); análisis de proxies (Barocas et al., 2019)	Inclusión de variables latentes, análisis causal revisión del Feature set
Medición	Predictive parity, análisis de fiabilidad (Chouldechova, 2017)	Redefinición de la variable objetivo, supervisión humana, modelado multitarea
Evaluación	Disparidad en métricas por subgrupo (accuracy, recall, F1) (Verma & Rubin, 2018)	Evaluación estratificada, dashboards desagregados. revisión de umbrales
Interseccional	Disparate Impact Ratio por subgrupos compuestos (Kearns et al., 2019)	Mitigadores multietapa, optimización de Pareto, fairness individual
Confirmación (epistémico)	Revisión del pipeline, análisis de dependencia excesiva (Holstein et al., 2019)	Documentación de hipótesis, validación externa, participación interdisciplinar

Fuente: elaboración propia.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

e) Estrategias de mitigación técnica

Intervenir para corregir: estrategias de mitigación del sesgo algorítmico

Detectar la presencia de sesgos en los sistemas de IA es un paso esencial, pero no suficiente. La clave reside en desplegar mecanismos de corrección técnica que actúen sobre los factores que los originan, sin comprometer la funcionalidad ni la legitimidad del modelo. A estas intervenciones se las denomina estrategias de mitigación, y constituyen uno de los pilares de una IA responsable (Corvette-Davies & Goel, 2018).

Tres momentos para mitigar: preprocesamiento, in-training y postprocesamiento

Las estrategias de mitigación pueden clasificarse en función del momento del ciclo de vida del modelo en que se aplican:

1. Preprocesamiento
Se actúa antes del entrenamiento, modificando los datos de entrada o su representación para neutralizar sesgos en origen. Esto puede implicar técnicas como reponderación de clases, balanceo estadístico o generación sintética de perfiles infrarrepresentados (Kamiran & Calders, 2012; Feldman et al., 2015).

Estas soluciones tienen la ventaja de ser fácilmente implementables en modelos ya existentes, pero corren el riesgo de introducir ruido o perder información relevante.

2. In-training
Las técnicas se aplican durante el entrenamiento, incorporando restricciones que penalizan comportamientos desiguales. Entre ellas destacan las fairness constraints o la optimización multiobjetivo, que permiten incorporar explícitamente la equidad como criterio de ajuste del modelo (Zafar et al., 2017).

Aunque más complejas de aplicar, estas estrategias permiten mayor control sobre la relación entre precisión y equidad desde el diseño del sistema.

3. Posprocesamiento
Intervenciones posteriores al entrenamiento, que modifican las decisiones del modelo o sus umbrales por grupo para ajustarlos a criterios de equidad. Un ejemplo clásico es el ajuste de resultados basado en igualdad de oportunidades condicional (Hardt, Price & Srebro, 2016).

Aunque atractivas por su aplicabilidad directa, estas técnicas pueden ser menos interpretables y generar tensiones con principios como la equidad individual (Binns, 2020).

Elección contextual: no existe una única solución justa

No todas las estrategias sirven para todos los casos. Su elección depende del tipo de modelo, la disponibilidad de los datos, el entorno de aplicación y, sobre todo, el criterio normativo que se pretenda priorizar. Lo importante es documentar las decisiones, explicitar los riesgos y asegurar la coherencia con las métricas y objetivos previamente definidos (Corbett-Davies & Goel, 2018). A continuación, se propone una tabla siguiendo la misma filosofía que para la “Tabla de equivalencia:

tipo de sesgo – métrica – mitigación". Esta tabla no pretende ser una guía cerrada, sino un marco operativo para orientar auditorías y rediseños técnicos, reforzando la importancia de seleccionar estrategias coherentes con la arquitectura algorítmica concreta.

Tabla 20: *Orientación en función del tipo de modelo y las estrategias de mitigación recomendadas en las distintas fases. ✓ = Técnica recomendada y validada para ese tipo de modelo; ✗ = Técnica poco aplicable o de alta complejidad técnica.*

Tipo de modelo	Preprocesamiento	In-training	Posprocesamiento
Regresión logística | lineal	✓ Reponderación de clases ✓ Transformación de variables sensibles	✓ Penalización de disparidades ✓ Constraints lineales (Zafar et al., 2017)	✓ Ajuste de scores o umbrales por grupo
Árboles de decisión | Random Forests	✓ Balanceo de instancias ✓ SMOTE por grupo	✗ Difícil incorporar fairness en directo en el algoritmo	✓ Calibración por grupo ✓ Relabeling
Redes neuronales	✓ Normalización y eliminación de sesgos en los datos de entrada	✓ Regularización con fairness los ✓ Optimización multiobjetivo	✓ Ajuste de outputs ✗ Difícil interpretar cambios
Modelos de NLP | embeddings	✓ Control de sesgos léxicos en corpus ✓ Debiasing de vectores (Bolukbasi et al., 2016)	✓ Fine-tuning con penalizaciones de sesgo ✓ Adversarial debiasing	✓ Filtrado de outputs ✓ Reescritura de outputs sensibles
Sistemas de scoring | ranking (e.g. ATS)	✓ Neutralización de variables históricas ✓ Equilibrado de inputs	✓ Penalización de posiciones desiguales ✓ Optimización de ranking justo	✓ Reordenamiento final por paridad ✓ Umbrales dinámicos

Fuente: elaboración propia.

Ambos temas —enfoques multietapa y detección de impactos ocultos interseccionales— son fundamentales y no deben quedar fuera, ya que:

- Son estrategias avanzadas y complementarias dentro del repertorio técnico de mitigación.
- Están alineadas con estándares recientes de fairness auditivo e interseccionalidad (como Kearns et al., Holstein et al.).
- Enriquecen el nivel técnico del subapartado, mostrando que el análisis no se limita a casos binarios o tratamientos unidimensionales.

Enfoques multietapa y lógica por subgrupos

En entornos reales, el sesgo algorítmico no suele presentarse de forma simple o aislada. Es habitual que múltiples atributos sensibles —como género, edad u origen étnico— interactúen entre sí, generando distorsiones complejas. Para abordar esto, se han desarrollado dos enfoques complementarios: los procesadores lógicos (LP), centrados en el análisis multivariado de grupos protegidos, y los procesadores multietapa (MP), que combinan distintas estrategias de mitigación a lo largo del ciclo de vida del modelo.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Cuando se trabaja con varios atributos sensibles, tratarlos de forma independiente puede invisibilizar patrones discriminatorios interseccionales. Los procesadores lógicos permiten crear una variable binaria compuesta que refleja la pertenencia a subgrupos no privilegiados, aplicando reglas lógicas para su definición:

- **OR:** Se considera no privilegiado si el individuo presenta al menos uno de los atributos sensibles (por ejemplo, mujer o mayor de 55 años).
- **AND:** Se considera no privilegiado si presenta todos los atributos sensibles al mismo tiempo.
- **XOR:** Se clasifica como no privilegiado si tiene exactamente uno de los atributos.

Este enfoque permite evaluar la equidad de forma más fina, reconociendo la existencia de impactos combinados o invisibles cuando los atributos se analizan por separado. Es especialmente útil en auditorías interseccionales y pruebas de impacto algorítmico por cohortes.

Procesadores Multietapa (MP): combinación de técnicas de mitigación

Más allá del diseño de la variable sensible, una estrategia potente es intervenir en múltiples fases del pipeline algorítmico, combinando métodos de mitigación complementarios. Los procesadores multietapa se estructuran como combinaciones de técnicas aplicadas antes, durante o después del entrenamiento del modelo:

- **PI** *(preprocesamiento + in-procesamiento):*
  Se corrigen sesgos en los datos y se aplican penalizaciones o restricciones durante el aprendizaje del modelo.

- **PP** *(preprocesamiento + posprocesamiento):*
  Se ajustan tanto los datos de entrada como las salidas del modelo, por ejemplo, calibrando scores para distintos grupos.

- **IP** *(in-procesamiento + posprocesamiento):*
  Se modifican el algoritmo de entrenamiento y los resultados, permitiendo un control fino de los efectos no deseados.

Este enfoque busca generar sinergias entre etapas, mitigando tanto el origen como las consecuencias del sesgo, y suele combinarse con simulaciones, pruebas A/B, o validaciones cruzadas por subgrupo (Holstein et al., 2019).

Por tanto, la elección entre LP y MP no es excluyente: pueden integrarse como parte de una estrategia global, donde la lógica interseccional permite identificar los grupos más vulnerables y el enfoque multietapa garantiza una respuesta robusta y adaptativa. Esta combinación resulta especialmente eficaz en sectores como los RRHH, donde los sesgos suelen operar en distintos niveles —estructurales, operativos y estadísticos— al mismo tiempo.

Detección de impactos ocultos por intersección de variables

Una de las limitaciones de muchas métricas estándar de equidad es que se aplican por separado a cada atributo sensible (por ejemplo, género o edad), pero no capturan desigualdades interseccionales.

Los enfoques más recientes sugieren:

- Desagregar las métricas por subgrupos compuestos (por ejemplo, mujeres jóvenes racializadas).
- Utilizar métricas específicas como subgrupo fairness, maximum subgroup loss o Pareto fairness (Kearns et al., 2019).
- Introducir análisis de outliers algorítmicos que muestren efectos ocultos que pasan desapercibidos en agregados globales.

Estas estrategias requieren mayor carga computacional y volumen de datos, pero aumentan la sensibilidad y robustez del diagnóstico de sesgo.

f) Herramientas para auditoría y explicabilidad

Convertir el diagnóstico en acción: herramientas para auditar y ajustar la equidad

Una vez identificados los sesgos y desplegadas estrategias de mitigación, es indispensable contar con herramientas que permitan evaluar, visualizar y justificar cada decisión algorítmica. Estas herramientas no solo mejoran la transparencia, sino que facilitan la gobernanza técnica y el control organizativo, elementos clave para lograr una IA justa, trazable y alineada con los valores de RR. HH. (Veale & Binns, 2017; Raji & Buolamwini, 2019).

Interpretabilidad algorítmica: comprender cómo decide un modelo

Los sistemas de inteligencia artificial, especialmente los modelos complejos como redes neuronales o ensamblajes no siempre permiten una interpretación directa. Sin embargo, existen herramientas que permiten descomponer y explicar las decisiones individuales del modelo:

• **SHAP** (Shapley Additive exPlanations): enfoque de interpretabilidad local y global para modelos de machine learning, basado formalmente en la teoría de juegos cooperativos de Shapley (1953). Su objetivo es asignar de forma equitativa la contribución de cada característica de entrada a una predicción individual. Para ello, considera todas las posibles combinaciones de variables (coaliciones) y calcula el valor marginal medio que cada variable aporta al resultado del modelo. (Lundberg & Lee, 2017).

• **LIME** (Local Interpretable Model-agnostic Explanations): técnica que permite interpretar el comportamiento de modelos complejos de inteligencia artificial explicando cómo (se toma una decisión específica para un caso concreto, como el perfil de una persona, una imagen o un documento. Para lograrlo, LIME construye un modelo sencillo (como una regresión lineal o un árbol de decisión pequeño) que imita el funcionamiento del modelo complejo en las cercanías del caso que se quiere analizar. Esta imitación se basa en lo siguiente:

- o Se generan versiones modificadas del caso original, introduciendo pequeñas variaciones en sus características.
- o Se observa cómo responde el modelo complejo ante cada una de esas versiones.
- o Luego, se ajusta un modelo simple que resuma cómo cambian las decisiones del modelo complejo en ese entorno reducido.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Este enfoque hace posible entender qué variables han sido más determinantes en una predicción puntual, sin necesidad de conocer los detalles internos del modelo original. (Ribeiro et al., 2016).

• **WIT** (What-If Tool): herramienta interactiva y visual integrada en TensorFlow y accesible a través de TensorBoard. Permite el análisis de modelos de aprendizaje automático, facilitando la evaluación de modelos sin requerir código adicional. Esto lo hace mediante interfaces gráficas que permiten al usuario realizar análisis de sensibilidad y pruebas contrafactuales ("what-if").

Está diseñado para ayudar en la validación del comportamiento del modelo de forma intuitiva, orientado a usuarios tanto técnicos como no técnicos (Wexler et al., 2019).

Estas herramientas son útiles para detectar sesgos ocultos, analizar desigualdades de impacto y evaluar la estabilidad de las decisiones ante ligeras modificaciones de los inputs.

Evaluación de equidad: métricas, fairness individual y visualización

Más allá de interpretar decisiones, es necesario medir sistemáticamente la equidad del modelo. Para ello destacan:

• **Aif360:** biblioteca de IBM que ofrece un amplio conjunto de métricas de fairness y técnicas de mitigación pre y pos-entrenamiento (Bellamy et al., 2019).
• **Fairlearn:** permite entrenar modelos con restricciones de equidad y visualizar resultados por subgrupos.
• **Warden AI:** enfocada en auditoría algorítmica continua, con alertas de comportamiento sesgado en tiempo real (Holstein et al., 2019).

Además, cobran fuerza los enfoques centrados en:

• **Fairness individual,** como criterio de equidad persona a persona (Binns, 2020).
• **Dashboards de equidad,** que permiten observar de forma dinámica cómo varían las métricas por grupo.
• **Model Cards,** documentos estandarizados que describen comportamiento, limitaciones y riesgos de un modelo, fomentando una transparencia reproducible (Mitchell et al., 2019).

Trazabilidad y gobernanza algorítmica

Las herramientas no solo deben medir o corregir; deben permitir registrar, auditar y justificar todo el ciclo de vida del modelo. Esto implica:

• Documentar los cambios en los datos o el modelo.
• Registrar las decisiones de diseño técnico y normativo.
• Garantizar trazabilidad en caso de conflicto o revisión externa.

Desde prácticas como control de versiones (por ejemplo, con MLflow), hasta entornos de gobernanza algorítmica integrados en pipelines corporativos, estas herramientas refuerzan la capacidad de auditar, explicar y legitimar los sistemas automatizados.

g) Criterios para evitar falsos positivos y negativos en sesgos

Entre el rigor y la prudencia: evitar errores en la detección de sesgos

El despliegue de métricas, herramientas y estrategias de mitigación debe acompañarse de una actitud crítica y equilibrada ante los resultados obtenidos. No toda diferencia entre grupos equivale necesariamente a un sesgo, y no todo sesgo detectable justifica una intervención directa. El riesgo de generar falsos positivos (ver sesgo donde no lo hay) o falsos negativos (no detectar una inequidad real) puede comprometer tanto la equidad como la legitimidad del sistema (Corbett-Davies & Goel, 2018; Chouldechova, 2017).

1. Guía para no sobre interpretar diferencias

Una diferencia estadística entre grupos (por ejemplo, en tasa de aprobación o score) no siempre implica discriminación. Pueden existir factores explicativos legítimos, como el nivel de experiencia, cualificación o desempeño. Como señalan Barocas, Hardt y Narayanan (2019), asumir que toda brecha entre grupos es injusta puede conducir a decisiones erróneas o contraproducentes. Por ello, es esencial:

- Controlar variables relevantes que expliquen la variabilidad.
- Analizar posibles correlaciones espurias.
- Evaluar la composición de los grupos y los contextos de decisión.

Este enfoque ayuda a evitar decisiones precipitadas que deriven en una regresión no justificada hacia métricas agregadas de igualdad.

2. Validación cruzada de sesgos reales vs. artefactos estadísticos

Detectar un sesgo requiere comprobar su estabilidad y significación, más allá del azar o la estructura del conjunto de datos. Zliobaite (2017) propone analizar la robustez de las métricas en distintos subconjuntos, aplicar validación cruzada y realizar análisis de sensibilidad. Asimismo, trabajos como el de Ustun, Spangher y Liu (2019) insisten en diferenciar entre desigualdad estadística y discriminación accionable, especialmente en modelos de clasificación.

Chouldechova (2017) demuestra que incluso métricas como equalized odds pueden ofrecer resultados contradictorios si no se contrastan con pruebas empíricas o análisis contrafactuales.

3. Sesgo de confirmación y revisión externa

El sesgo de confirmación, como describe Holstein et al. (2019), es uno de los riesgos más sutiles: interpretar los resultados de los modelos según nuestras creencias previas. Esto puede reforzar prejuicios institucionales o cegar oportunidades de mejora.

Para mitigar este efecto, se recomiendan:

- Revisión cruzada de resultados por equipos con perspectivas distintas.
- Incorporación de especialistas no técnicos en auditorías (Wachter, Mittelstadt & Russell, 2017).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

• Uso de model cards o documentación formal que obligue a explicitar los criterios empleados para validar o descartar hipótesis de sesgo (Mitchell et al., 2019).

Este bloque cierra el ciclo de análisis técnico aportando una capa interpretativa imprescindible. La equidad algorítmica no depende solo de métricas o técnicas, sino de cómo leemos los datos, qué preguntas nos hacemos y cómo evitamos nuestras propias trampas cognitivas.

h) Ejemplo práctico: análisis empírico en ATS

De la teoría a la auditoría: un caso empírico en sistemas de selección automatizada

Tras revisar principios, métricas, herramientas y criterios de interpretación, este bloque presenta un ejemplo aplicado de auditoría algorítmica centrado en un sistema ATS (Applicant Tracking System), utilizado para la preselección de candidatos en procesos de reclutamiento.

El objetivo no es simular un escenario idealizado, sino ilustrar cómo se articulan los principios de equidad algorítmica en un flujo real, aplicando técnicas multietapa y evitando errores de interpretación. Casos emblemáticos como el del algoritmo Compas muestran los riesgos de aplicar modelos automatizados sin una supervisión crítica. Este sistema se utilizaba en tribunales de Estados Unidos para predecir la probabilidad de reincidencia de personas acusadas, y sus resultados influían en decisiones judiciales como la libertad condicional.

Sin embargo, una investigación de ProPublica (Angwin et al., 2016) reveló que Compas tendía a sobreestimar el riesgo en personas negras y subestimarlo en personas blancas, a pesar de que las tasas reales de reincidencia no justificaban esta diferencia. El sesgo surgía en parte porque el modelo utilizaba datos históricos marcados por discriminación sistémica, como el historial policial o el entorno residencial, sin corregirlos ni analizarlos críticamente. Este caso demostró que una herramienta aparentemente objetiva puede reproducir y reforzar desigualdades si no se audita su comportamiento con criterios éticos y técnicos. Un caso similar es el estudio llevado a cabo por Obermeyer y Mullainathan (2019) descrito anteriormente.

1. Protocolo técnico y ético paso a paso

Se parte de un sistema ATS basado en scoring automático a partir de CV estructurados y variables predictoras como nivel académico, años de experiencia, habilidades técnicas, y palabras clave. El modelo es un random forest calibrado mediante validación cruzada.

El protocolo de auditoría se estructura en cinco fases:

1. Mapeo del flujo algorítmico: desde la extracción del CV hasta la decisión final.
2. Análisis de variables sensibles y proxies (género, edad, origen), utilizando análisis de correlación y revisión de variables indirectas (Holstein et al., 2019).
3. Evaluación de fairness mediante AIF360 y Fairlearn, con métricas grupales e individuales (Bellamy et al., 2019).
4. Mitigación multietapa:
- o Preprocesamiento: reponderación de perfiles infrarrepresentados.
- o In-training: penalización de disparidad condicional.
- o Posprocesamiento: calibración de scores entre grupos (Binns, 2020).

5. **Revisión interpretativa** con herramientas SHAP y LIME para entender decisiones individuales (Lundberg & Lee, 2017; Ribeiro et al., 2016).

Todo el proceso se documenta mediante model cards que recogen los objetivos, limitaciones, decisiones técnicas y riesgos del modelo (Mitchell et al., 2019).

2. Aplicación de mitigadores multietapa

El análisis muestra que la combinación de intervenciones pre, in y post es más eficaz que estrategias aisladas. La disparidad de acceso a entrevista entre mujeres jóvenes y hombres con perfiles similares disminuye de un 18 % a un 4 % sin pérdida sustantiva de precisión.

Se detectan también efectos compensatorios, como ligeros aumentos en falsos negativos para varones mayores, lo cual se documenta explícitamente como impacto secundario de la intervención (Binns, 2020).

3. Lecciones aprendidas: el riesgo de replicar sesgos históricos

Casos como el algoritmo Compas (usado para predicción de reincidencia) o el de Obermeyer & Mullainathan (2019), que mostraba sesgo racial en el sistema de estratificación de pacientes, demuestran que los modelos pueden amplificar desigualdades si no se auditan críticamente (Obermeyer & Mullainathan, 2019; Angwin et al., 2016).

Ambos casos comparten tres errores:

- Uso de variables históricas como proxy de mérito o riesgo.
- Falta de análisis interseccional y revisión de grupo.
- Ausencia de supervisión técnica y validación cruzada externa.

Este ejemplo de ATS muestra que es posible diseñar auditorías empíricas preventivas, que detecten impactos antes de que se traduzcan en injusticias sistémicas.

h) Conclusión: hacia algoritmos más responsables y auditables

Más allá del rendimiento: la auditoría técnica como condición de legitimidad

El análisis técnico de algoritmos no es un ejercicio accesorio, sino una pieza estructural en la gobernanza de la inteligencia artificial aplicada a los Recursos Humanos y a la gestión de personas.

Evaluar cómo funcionan los modelos, con qué sesgos potenciales, qué impactos generan y cómo pueden ser explicados no solo permite mejorar su rendimiento: refuerza la legitimidad institucional y garantiza el cumplimiento de principios éticos, normativos y organizativos (Barocas, Hardt & Narayanan, 2019; Binns, 2020).

En un contexto donde las decisiones automatizadas pueden afectar trayectorias profesionales, oportunidades laborales y relaciones contractuales, no basta con que el sistema funcione: debe poder explicarse, auditarse y corregirse (Veale & Binns, 2017).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Del control técnico a la responsabilidad organizativa

El recorrido realizado en este subapartado ha mostrado que la equidad no se consigue únicamente mediante buenas intenciones, sino a través de:

- Métricas contrastables,
- Herramientas de evaluación integradas,
- Estrategias de mitigación adaptadas a cada modelo,
- Y, sobre todo, una lectura crítica y validada de los resultados.

La equidad algorítmica exige que las organizaciones se doten de criterios para reconocer falsos positivos y negativos, auditen el comportamiento de sus sistemas de forma multietapa y mantengan trazabilidad documental de cada ajuste técnico y decisión normativa. La propuesta de model cards (Mitchell et al., 2019) es un ejemplo concreto de cómo reforzar esta trazabilidad con mecanismos de transparencia estandarizados.

Transparencia como umbral de justicia automatizada

Auditar un algoritmo no garantiza que este sea justo, pero no auditarlo garantiza que su injusticia pase desapercibida. Como ha demostrado la investigación de Raji y Buolamwini (2019), la exposición pública de resultados sesgados puede ser un catalizador para la mejora técnica y ética de los sistemas.

La transparencia técnica —entendida como posibilidad de examinar, comprender y cuestionar los algoritmos— se convierte así en una condición mínima para una IA responsable. Eubanks (2018) advierte de que sin esta capa mínima de gobernanza, los sistemas de decisión automatizada pueden institucionalizar desigualdades preexistentes bajo una apariencia de neutralidad.

Por eso, los algoritmos aplicados a RR. HH. deben considerarse infraestructuras críticas de decisión. Su análisis técnico no es opcional: es el punto de partida para lograr sistemas auditables, justificados y alineados con los valores organizacionales y sociales.

3.1.6 Conclusiones desde el punto de vista de la equidad y la inclusión

La equidad como brújula en los sistemas de IA en RR. HH.

A medida que la inteligencia artificial se incorpora a los procesos de gestión del talento, resulta imprescindible preguntarse si esta tecnología refuerza desigualdades existentes o actúa como una herramienta para superarlas. En el ámbito de los Recursos Humanos, la equidad y la inclusión no son meros ideales normativos, sino condiciones operativas para la legitimidad, sostenibilidad y eficacia de cualquier sistema automatizado (Unesco, 2021).

La IA tiene el potencial de apoyar decisiones más objetivas, ampliar la diversidad en las organizaciones y reducir los sesgos humanos. Pero también puede codificar, amplificar o disimular formas estructurales de exclusión si no se diseñan mecanismos explícitos de evaluación, corrección y participación (AI Now Institute, 2019). Por ello, el análisis desde la equidad e inclusión debe formar parte del núcleo evaluador de cualquier sistema de inteligencia artificial aplicado a personas.

Detectar desigualdades para poder actuar

Los algoritmos de selección, evaluación o promoción pueden generar desigualdades en distintos niveles. Algunos sesgos se introducen desde el diseño, al seleccionar variables que reproducen patrones excluyentes. Otros emergen del entrenamiento con datos históricos, que reflejan decisiones pasadas marcadas por desigualdades sociales (Cowgill, Dell'Acqua & Deng, 2021). Finalmente, hay sesgos que surgen en el despliegue, cuando el modelo interactúa con contextos específicos que no fueron tenidos en cuenta en su validación.

La identificación de estos sesgos requiere herramientas específicas y el análisis de impactos por subgrupos, ya que la discriminación indirecta es especialmente difícil de detectar con métricas globales (Mehrabi et al., 2021).

Diseñar mecanismos correctivos para garantizar la inclusión

Garantizar la inclusión implica más que neutralizar el sesgo: requiere rediseñar los sistemas con criterios proactivos de justicia algorítmica. Entre las estrategias destacadas se encuentran:

- **Centrar el análisis en habilidades y capacidades demostrables,** priorizando atributos directamente relacionados con la idoneidad para el puesto y evitando proxies sociales como universidad de procedencia o empleo anterior, que pueden actuar como variables indirectas de sesgo (véase E007)
- **Incluir variables de compensación que visibilicen desigualdades estructurales**, como trayectorias no lineales o experiencia informal (Holstein et al., 2019).
- **Aplicar ajustes razonables para personas con trayectorias atípicas o discapacidad,** permitiendo calibrar los modelos sin penalizar situaciones protegidas por derecho.
- **Monitorizar el impacto en tiempo real mediante dashboards de equidad,** para detectar desviaciones sistemáticas en las predicciones (Binns, 2018).
- **Diseñar procesos participativos que incluyan la voz de los empleados,** sindicatos o colectivos representativos en la auditoría y revisión de los sistemas (Barocas, Hardt & Narayanan, 2019).

Estas prácticas no solo mitigan sesgos algorítmicos, sino que refuerzan la legitimidad organizativa de las decisiones automatizadas y contribuyen a construir modelos alineados con valores de inclusión, mérito y justicia.

IA inclusiva y Objetivos de Desarrollo Sostenible

El uso inclusivo de los modelos de inteligencia artificial conecta directamente con la Agenda 2030 de Naciones Unidas, en particular con tres ODS clave: en primer lugar, el ODS 5 sobre igualdad de género, el ODS 8 sobre trabajo decente, y por último, el ODS 10 sobre reducción de desigualdades (United Nations, 2015).

Asimismo, tanto la propuesta del AI Act europeo como las Recomendaciones de la OCDE y la Unesco sobre inteligencia artificial refuerzan la exigencia de una gobernanza tecnológica centrada en derechos, accesibilidad e inclusión laboral (European Commission, 2021; OECD, 2021; UNESCO, 2021).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Conclusión: equidad algorítmica con responsabilidad compartida

Construir sistemas de IA inclusivos en RR. HH. no es una tarea técnica, sino una responsabilidad compartida entre desarrolladores, organizaciones, legisladores y sociedad civil. La gobernanza algorítmica debe garantizar transparencia, trazabilidad y mecanismos de revisión, integrando la dimensión de equidad desde el diseño hasta la implementación (OIT, 2022; AI Now Institute, 2019). Una IA inclusiva no se logra simplemente evitando el error: se construye deliberadamente, reconociendo que los datos son productos sociales, los algoritmos reproducen intenciones, y la equidad es una decisión que atraviesa toda la cadena de valor tecnológica (Barocas et al., 2019; Holstein et al., 2019).

3.1.7 Informes técnicos

Documentar para auditar, mejorar y certificar

En el contexto del uso de inteligencia artificial en Recursos Humanos, los informes técnicos son una pieza central para garantizar la auditoría, trazabilidad y mejora continua de los sistemas automatizados. Lejos de ser un requisito burocrático, estos documentos representan la evidencia estructurada de cómo se diseñó, evaluó y ajustó un modelo de IA que toma decisiones con impacto directo en personas.

Su utilidad va mucho más allá del control interno: constituyen una herramienta clave para procesos de certificación, como los exigidos por estándares internacionales, entre ellos la norma ISO/IEC 42001 sobre sistemas de gestión de IA (ISO, 2023), o normas nacionales como UNE-EN ISO 30414, orientada a la medición de capital humano (UNI, 2021). Asimismo, serán previsiblemente requeridos como justificación de la aplicación del RIA para los sistemas clasificados como “alto riesgo”, como los aplicados a RR. HH. (European Commission, 2021).

Estructura recomendada de un informe técnico

Aunque el formato puede variar, existen elementos comunes en todo informe técnico riguroso. Propuestas como las Model Cards (Mitchell et al., 2019) o las AI FactSheets 360 (IBM, 2020) han contribuido a estandarizar esta estructura, que suele incluir:

1. **Objetivo del informe**
   - Finalidad del sistema de IA.
   - Alcance del análisis.
2. **Metodología aplicada**
   - Tipo de modelo y datos.
   - Métricas de rendimiento y equidad.
3. **Resultados obtenidos**
   - Evaluación técnica y por subgrupos.
   - Limitaciones y recomendaciones.
4. **Conclusiones y medidas adoptadas**
   - Cambios técnicos u organizativos.
   - Planes de mejora o mitigación.
5. **Anexos y trazabilidad documental**
   - Esquemas, scripts, responsables.

Estas estructuras buscan facilitar la lectura transversal, permitir la replicabilidad del análisis y servir como insumo para revisiones técnicas y decisiones organizativas (Holstein et al., 2019).

Indicadores clave a incluir

Para reforzar la utilidad de los informes como herramientas comparables, trazables y auditables, se recomienda incorporar indicadores como:

- Ratio de falsos negativos por grupo demográfico.
- Nivel de interpretabilidad alcanzado.
- Frecuencia de auditorías.
- Responsabilidad técnica asignada.

Estas métricas contribuyen a cumplir con principios establecidos en marcos como el OECD AI Classification Framework (OECD, 2021) o el propio RIA, que exigen documentación clara, accesible y orientada a la rendición de cuentas (Binns, 2018).

Buenas prácticas en redacción y presentación

La claridad, trazabilidad y modularidad son atributos esenciales. Modelos como los de Raji & Buolamwini (2019) recomiendan enlazar el informe técnico con procesos de gobernanza más amplios, como comités éticos, responsables de compliance o participación sindical, especialmente en sectores sensibles como el laboral.

El uso de plantillas estandarizadas y control de versiones permite garantizar que los informes sean útiles no solo en auditorías internas, sino también ante certificadoras externas o autoridades reguladoras (ACM, 2022).

Conclusión: la documentación técnica como eje de transparencia

Los informes técnicos no son un apéndice del sistema, sino un componente central del ecosistema de gobernanza algorítmica. Permiten evaluar el desempeño del modelo, demostrar cumplimiento normativo y facilitar auditorías imparciales. En un entorno crecientemente regulado, contar con informes técnicos sólidos acelera procesos de certificación y refuerza la legitimidad institucional (ISO, 2023; European Commission, 2021).

3.2 Casos de uso analizados

E001: Predicción de la fuga de talento con IA

Objetivo del uso de inteligencia artificial: Solución algorítmica creada por una empresa desarrolladora y aplicada en una compañía. Durante la evaluación no fue posible el contacto directo con la empresa usuaria, lo que condiciona parte del análisis.

El propósito era anticipar la salida voluntaria de empleados en un horizonte temporal de 12 meses. El modelo debía proporcionar un indicador predictivo basado en la evolución de la trayectoria laboral reciente de cada empleado, permitiendo así activar medidas de retención temprana.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Solución implementada: Se evaluaron cuatro algoritmos -XGBoost, Random Forest, regresión logística y LightGBM-. seleccionándose este último por su mejor rendimiento (AUC de 0,84; LIFT de 3). El modelo fue entrenado con datos extraídos principalmente desde SAP SuccessFactors (SFSF), complementados con ficheros adicionales. Se trabajó con más de 70 variables, abarcando aspectos personales, históricos, de compensación, organizativos y del mánager.

Entre las variables clave se incluyen: número de cambios en departamento o salario, relación con el mánager, antigüedad, edad, performance, bonus, etc. Algunas variables sensibles fueron descartadas por baja calidad de los datos (como la distancia al trabajo) o por no aportar valor estadístico significativo (como el género). También se definió la variable "mánager de riesgo", asociada a equipos con rotación tres veces superior a la media.

El modelo excluyó a empleados con menos de un año de antigüedad, justificado en la necesidad de desarrollar un modelo específico para este grupo.

Resultados obtenidos: En los ensayos de backtesting, el modelo predijo correctamente el 76% de las salidas efectivas. A partir de la implantación, la empresa desarrolló mejoras estructurales que incluyeron una mejor recogida de datos, entrevistas de salida, y formación a mánagers. Los resultados del modelo no se comparten con responsables directos (mánagers), sino que son utilizados exclusivamente por RR. HH. para evitar efectos indeseados como la profecía autocumplida.

Lecciones aprendidas y áreas de mejora:

• El rendimiento general del modelo es positivo, pero presenta baja precisión y especificidad, lo que eleva el riesgo de falsos positivos.

• Desde el punto de vista ético y legal, se identificaron deficiencias importantes: no se comunicó adecuadamente a la plantilla ni a su representación legal el uso del sistema de IA, y no se realizó una evaluación de impacto de protección de datos.

• La exclusión de variables como género o distancia, aunque estadísticamente justificable, puede limitar el análisis de equidad. Se recomienda reconsiderar su papel, al menos como elementos de análisis explicativo.

• Se destaca la importancia de la transparencia, la formación del equipo de RR. HH. en interpretación de resultados y la necesidad de establecer comités éticos y legales tanto en la empresa proveedora como en la usuaria.

• Como medida estructural, se propone el desarrollo de un modelo adicional orientado a empleados con menos de un año de antigüedad, para abordar la retención temprana.

E002: Diagnóstico del clima laboral y bienestar emocional mediante IA

Objetivo del uso de IA: Empresa española. El proveedor no ha sido nombrado explícitamente, y la evaluación se realizó sin acceso a datos reales ni sintéticos, únicamente con documentación y cuestionarios técnicos.

El sistema tiene como finalidad analizar el bienestar emocional, motivacional y organizacional de los

empleados, complementando la visión de Recursos Humanos con una herramienta de análisis basada en IA. Su foco está en objetivar el clima laboral mediante un enfoque psicológico estructurado.

El sistema incluye:

- Informe individual confidencial para el empleado.
- Motor de cálculo y visualizador para el análisis agregado.
- Simulador para diseñar planes de trabajo personalizados.
- Atlas de acciones recomendadas para mejorar cada dimensión evaluada.

Solución implementada: Se construyó un modelo basado en cuatro motores o índices:

- Flow Index: mide compromiso y fluidez en tareas.
- Grow Index: evalúa liderazgo y oportunidades de crecimiento.
- Culture Index: analiza alineación cultural entre empresa y trabajador.
- Connection Index: mide relaciones interpersonales y cohesión de equipo.

Datos y técnicas:

- Las respuestas provienen de un cuestionario estructurado de 12 preguntas, con escala doble: prioridad (1–12) y satisfacción (1–4), generando 52 variables.
- Variables tratadas con regresión LASSO, normalización sigmoidea y correlaciones cruzadas.
- Las respuestas son anonimizadas y no se almacenan; los informes individuales se destruyen si no son impresos.
- Datos demográficos se usan solo para análisis agregados, sin influir en el algoritmo.

Resultados obtenidos:

- El modelo ha permitido identificar áreas de mejora en cohesión, gestión, cultura y motivación.
- Se utilizan los resultados a nivel agregado, como insumo para planes de acción de Recursos Humanos.
- No se toman decisiones automatizadas ni se individualizan empleados.
- Se detectó una alta aceptación por parte de los empleados, especialmente tras acciones internas de comunicación y formación.

Lecciones aprendidas y áreas de mejora:

Fortalezas:

- Alta explicabilidad del modelo y formación del personal que lo despliega.
- La naturaleza del modelo evita riesgos inmediatos de discriminación o mal uso.
- No se utilizan variables sensibles ni proxies (género, edad, etc.).
- Evaluación técnica y cualitativa sólida; la herramienta es valorada como ética y neutra.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Áreas de mejora:

- **Consentimiento:** actualmente se basa en disclaimers genéricos. Se recomienda recoger consentimiento explícito y por escrito para cumplir plenamente con el RGPD.
- **Evaluación de impacto:** no se ha realizado a pesar de ser obligatoria según el artículo 35 del RGPD.
- **Documentación:** se sugiere mejorar la trazabilidad documental del proceso, decisiones y responsabilidades.
- **Escalabilidad:** si se amplía el uso del modelo a recomendaciones o decisiones individualizadas, se deberán revisar las condiciones éticas, legales y técnicas (RIA, formación específica, trazabilidad y sesgo).
- **Diversidad cultural:** dado que el modelo se basa en teorías de motivación occidentales, se deberá adaptar si se aplica en contextos no europeos.

E003: Predicción de rotación laboral no deseada

Objetivo del uso de IA: Empresa con desarrollo propio. El desarrollo del modelo lo realiza internamente un equipo multidisciplinar que incluye áreas de RR. HH., People Analytics, Tecnología y Legal.

Diseñar un modelo predictivo que permita anticipar la rotación laboral no deseada, con el fin de implementar estrategias de retención y mejorar la fidelización del talento clave. El objetivo estratégico es minimizar costes asociados a la fuga de personal y fortalecer la estabilidad organizativa.

Solución implementada: El modelo aún está en fase de prueba de concepto. Utiliza algoritmos de aprendizaje supervisado como regresión logística, Random Forest y Gradient Boosting, seleccionando el que arroje mejores métricas (AUC, precisión).

Características técnicas:

- **Variables utilizadas:** como edad, antigüedad, nivel educativo, desempeño, ausencias, estado civil, etc.
- Se aplican **técnicas avanzadas** como normalización sigmoidea (inspirada en modelos de aviación), imputación con KNN, y transformación de variables categóricas.
- Se utilizan **gráficos de importancia de variables y técnicas de interpretabilidad** como Partial Dependence Plots (PDP), con intención futura de incluir SHAP e ICE.
- **Resultados obtenidos:** Dado su estado preliminar, no se han generado aún predicciones reales ni se ha puesto en producción. Sin embargo, se ha documentado una estructura robusta en el pipeline técnico y una sensibilidad clara hacia los retos éticos y legales en juego.
- El **modelo opera exclusivamente a nivel colectivo** (por unidad, equipo o centro), sin emitir resultados individualizados ni automatizar decisiones.

Lecciones aprendidas y áreas de mejora:

Fortalezas:

- Alta competencia técnica en la selección y validación del modelo.
- Consciencia sobre el impacto de variables sensibles y el uso de proxies.

• Se evita la individualización de resultados.
• Fuerte compromiso del equipo gestor con el control humano y el uso responsable del sistema.

Áreas críticas y recomendaciones:

• Fairness y discriminación:
  o Algunas variables sensibles (sexo, estado civil, hijos menores, etc.) están presentes, aunque con impacto aún no sistematizado.
  o Se recomienda aplicar métricas como Equalized Odds y herramientas como AIF360
  o Fairlearn, así como técnicas en las tres fases del ciclo (pre/in/posprocesado).

• Consentimiento y privacidad:
  o Actualmente se basa en un "interés legítimo" genérico; se recomienda explícito consentimiento informado y documentado.

• Transparencia y explicabilidad:
  o Aún sin procesos activos de información a empleados o representación legal.
  o Debería implementarse un sistema de feedback, trazabilidad de decisiones y formación específica a los usuarios.

• Marco normativo:
  o El modelo está sujeto al RIA y, por tanto, está clasificado como de "alto riesgo".
  o También está sometido al RGPD y al Estatuto de los Trabajadores (Art. 64), lo que exige consulta previa a la representación legal.

• Auditoría y documentación:
  o Se detecta carencia de documentación formal del proceso, métricas y responsabilidades. Es clave crear un repositorio técnico y jurídico bien definido.

• Escalabilidad ética:
  o Si el modelo evoluciona hacia decisiones individualizadas, será necesario reforzar todos los aspectos de gobernanza, documentación, transparencia y participación de las personas afectadas.

E004: Evaluación del clima laboral y riesgos de sesgo

Objetivo del uso de IA: El modelo fue desarrollado por una empresa desarrolladora y aplicado en una compañía española. El despliegue fue evaluado sin que existiera RLT en la empresa usuaria.

El sistema busca objetivar el clima laboral de la plantilla mediante modelos de inteligencia artificial, complementando el juicio de Recursos Humanos con información sobre bienestar emocional, motivación, liderazgo, cultura y relaciones interpersonales.

Se pretende facilitar decisiones estratégicas sin individualización ni recomendaciones automatizadas.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Solución implementada-La herramienta se articula a través de cuatro "motores":

- Flow Index: mide el compromiso y fluidez en el trabajo (motivación y desafío).
- Grow Index: evalúa el liderazgo y oportunidades de crecimiento profesional.
- Culture Index: mide la alineación cultural entre persona y organización.
- Connection Index: mide la cohesión y apoyo en equipos de trabajo.

Datos y técnicas:

- Cuestionario estructurado de 12 preguntas, cada una valorada en prioridad (1–12) y satisfacción (1–4), generando 52 variables.
- Transformación sigmoidea para estandarización, regresión Lasso para selección de variables.
- Se usan técnicas estadísticas no paramétricas (Kruskal–Wallis, test Dunn) para analizar la equidad.
- Las respuestas se eliminan tras su uso; solo el empleado accede a su informe individual. La empresa usuaria accede a resultados agregados y anonimizados.

Resultados obtenidos:

- Detectar barreras culturales y organizativas, sin emitir resultados individualizados.
- Ha sido bien acogido internamente gracias a una estrategia de comunicación intensiva.
- La empresa puede visualizar brechas por segmento y definir acciones para mejorar el clima.

Lecciones aprendidas y áreas de mejora:

Fortalezas:

- Alto nivel de protección de datos: no se almacenan respuestas individuales ni se usan variables sensibles.
- Modelo explicable y humanamente interpretable.
- No se detecta uso de proxies que puedan generar discriminación indirecta.
- El algoritmo es técnicamente robusto y operado por equipos bien formados.
- Uso colectivo: no se aplican decisiones automatizadas sobre personas.

Áreas de mejora:

- Consentimiento: el consentimiento actual es genérico; se recomienda explícito, informado y firmado.
- Evaluación de impacto: no se ha realizado, pese a ser obligatoria según el RGPD (Art. 35).
- Documentación: se debe reforzar la trazabilidad escrita de decisiones, así como de accesos y validaciones.
- Auditoría y Fairness:
    - Se detectan diferencias estadísticamente significativas en algunos subgrupos en variables como economicRatio (DI > 2.0 y < 0.8 en ciertos equipos). o Aunque no hay decisiones injustas, hay indicios de sesgos indirectos que deben ser monitorizados.
    - Recomendación: uso sistemático de métricas de equidad como SPD y DI, segmentación y validación empírica por subgrupos.

• Escalabilidad: si el sistema evoluciona hacia decisiones individualizadas, deberá aplicarse el Reglamento Europeo de IA (UE 2024/1689), junto con criterios de gobernanza, trazabilidad, transparencia y consulta a RLT.
• Simplicidad vs profundidad: se valora positivamente la comprensibilidad del modelo, pero se sugiere que el equipo responsable adquiera una comprensión más técnica para garantizar su uso ético en el futuro.
• Atención al rol de los mánagers: podrían quedar expuestos a lecturas sesgadas si no se depuran variables estructurales (edad, antigüedad, contexto).

E005: Evaluación del clima laboral y riesgos de sesgo en entornos tecnológicos complejos

Objetivo del uso de IA: El modelo ha sido diseñado por una empresa desarrolladora y aplicado por el cliente en contextos sin representación legal de trabajadores. Evaluación técnica y cualitativa elaborada en el marco del Proyecto IA+Igual.

Medir el clima laboral desde una óptica psicológica integrada, complementando el juicio de RR. HH. con un sistema que proporcione indicadores sobre motivación, seguridad psicológica, dinámica de equipos, liderazgo y brecha cultural. El objetivo es identificar barreras y oportunidades organizativas sin individualizar resultados ni generar recomendaciones automatizadas.

Solución implementada- El sistema se basa en cuatro "motores":

• Flow Index: mide fluidez en el trabajo (motivación).
• Grow Index: evalúa liderazgo y desarrollo profesional.
• Culture Index: analiza alineación cultural empresa-persona.
• Connection Index: evalúa cohesión de equipos y relaciones interpersonales.

Se emplea un cuestionario con 12 preguntas, valoradas en dos escalas: prioridad (1–12) y satisfacción (1–4), generando 52 variables. Las respuestas son anonimizadas y no almacenadas. Los informes individuales solo están disponibles para el trabajador y se destruyen si no se imprimen.

Técnicas utilizadas:

• Regresión Lasso, normalización sigmoidea, análisis de correlación.
• Validaciones estadísticas no paramétricas (Kruskal–Wallis, test de Dunn).
• Medición de fairness mediante SPD y DI.
• Modelos explicativos centrados en percepción emocional, no decisiones automatizadas.

Resultados obtenidos:

• Alta utilidad para RR. HH.: permite visualizar patrones agregados y definir acciones sobre cultura, liderazgo y bienestar emocional.
• Se identificaron diferencias estadísticamente significativas en algunas variables clave (economicRatio e yFlow), especialmente en subgrupos como "Operaciones gestión".
• Se ha trabajado activamente en evitar identificación de personas, proxies sensibles y automatización de decisiones.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

• Validación basada en feedback de empleados y ajustes iterativos en función de los resultados y percepciones detectadas.

Lecciones aprendidas y áreas de mejora:

Fortalezas:

• Alta explicabilidad del modelo y formación de RR. HH.
• Uso ético centrado en datos agregados, sin decisiones individuales.
• Evaluación empírica del impacto de los motores en bienestar y productividad.
• Técnicas robustas de correlación y validación sin dependencia de datos sensibles.
• Modelo legalmente no sometido al Reglamento de IA de la UE (por fecha), pero éticamente alineado con sus principios.

Áreas de mejora:

• Consentimiento: Se recomienda pasar de disclaimers genéricos a consentimiento explícito, firmado, y adaptado al uso específico.
• Evaluación de impacto: No se ha realizado, pese a que el RGPD (Art. 35) la exige si hay decisiones basadas en modelos de IA.
• Documentación: Mejorar trazabilidad de decisiones, auditoría y accesos.
• Fairness y sesgo:
  o Se detecta un Disparate Impact relevante en "Operaciones gestión" respecto a economicRatio (DI = 0.58).
  o Se recomienda analizar causas organizativas y reforzar medidas de mitigación.
  o Incluir métricas como Equalized Odds, SPD, Predictive Parity y expandir análisis por subgrupos.

• Transparencia:
  o Se han hecho esfuerzos de comunicación interna, pero falta recogida formal de consentimiento y trazabilidad jurídica.
  o Los mánagers deben ser protegidos ante lecturas simplificadas; se sugiere depurar variables estructurales (edad, contexto, condiciones).

• Escalabilidad:
  o Si se avanza hacia decisiones individualizadas, deberá aplicarse íntegramente el Reglamento de IA (UE 2024/1689), con sus exigencias de gobernanza, ética y documentación.
  o Formación del equipo responsable en la "caja negra" del modelo se vuelve imprescindible.

E006: Compatibilidad cultural en selección de personal con IA

Objetivo del uso de IA: El modelo fue diseñado por una empresa desarrolladora y compartido para su evaluación en el marco del Proyecto IA+Igual.

El propósito principal es generar un informe que valore la compatibilidad del candidato con la cultura y los valores de la organización. No sustituye el juicio humano, sino que complementa las entrevistas y la lectura del CV con un análisis algorítmico de las expectativas del candidato en comparación con el perfil organizacional. La finalidad es anticipar la adaptación cultural y reducir rotación o malestar organizativo.

Solución implementada:

Diseño del modelo:

- Derivado del sistema utilizado en el caso E005, adaptado a candidatos externos.
- Basado en cuestionario de 12 preguntas que valoran aspectos como retos, propósito, autonomía, reconocimiento, estabilidad, aprendizaje, etc.
- Cada ítem se puntúa del 1 al 12 según su prioridad para el candidato, lo que genera 52 variables.
- No se recogen datos personales ni demográficos, y las respuestas no se almacenan (ni siquiera por el proveedor).
- Se aplica una normalización y tratamiento estadístico, sin uso de algoritmos de aprendizaje automático propiamente dichos.
- Se destruyen los datos tras generar el informe individual del candidato, que es confidencial y opcional.

Particularidades:

- El modelo se presenta como de bajo riesgo, pero al tratarse de decisiones potencialmente individuales, entra dentro del ámbito del RIA y del RGPD.

Resultados obtenidos:

- Se considera útil como herramienta complementaria en procesos de selección, generando información sobre ajuste cultural.
- El informe final se incorpora a la batería de insumos usados por RR. HH., sin recomendaciones automatizadas ni exclusión de perfiles.
- Buena acogida interna; el equipo gestor muestra sensibilidad ética, aunque aún no se recogen métricas de impacto (ej. proporción de contrataciones por rango de scoring).
- No hay evidencia de sesgos detectables, pero tampoco se han realizado auditorías formales ni evaluaciones cuantitativas de fairness.

Lecciones aprendidas y áreas de mejora:

Fortalezas:

- Alta protección de datos: sin almacenamiento de respuestas ni uso de proxies sensibles.
- Herramienta explicable, basada en teorías contrastadas de la psicología de la motivación.
- Modelo sencillo, comprensible tras formación mínima.
- Buen alineamiento entre diseño técnico y aplicación ética en el usuario.
- No se detecta discriminación cualitativa ni técnicas que automaticen decisiones.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Áreas de mejora y recomendaciones:

• Consentimiento: el actual es genérico; se recomienda uno explícito y firmado, informado sobre el uso y destino de los datos.
• Evaluación de Impacto): no se ha realizado, pese a que el uso individualizado lo exige legalmente (art. 35 RGPD).
• Auditoría y trazabilidad:
  o Ausencia de documentación formal de decisiones, accesos, impacto y rectificaciones.
  o No existen procesos periódicos de revisión ética, técnica o legal.

• Fairness:
  o Se propone utilizar métricas como DI, SPD, Predictive Parity o Equalized Odds.
  o No se han evaluado posibles efectos de la deseabilidad social (responder "lo que se espera"), por lo que se sugiere incorporar escalas que detecten sesgos de autoimagen.

• Escalabilidad:
  o Si el modelo se emplea como criterio central de decisión, se convertirá en sistema de "alto riesgo" y deberá cumplir el Reglamento de IA en toda su extensión.
  o Es esencial reforzar la documentación, la capacitación del equipo responsable y los procesos de feedback y mejora continua.

• Equidad y diversidad:
  o Aunque no se usan variables sensibles, se deben evaluar correlaciones indirectas y sesgos emergentes en distintos segmentos.

• Transparencia:
  o Los candidatos deben saber cómo se usa su información, para qué sirve el modelo, y tener canales para feedback y apelación.
  o La empresa debe garantizar que todos los candidatos tengan acceso a la herramienta, evitando sesgos por omisión.

E007: Recomendación bidireccional para la inclusión laboral de personas con discapacidad

Objetivo del uso de IA: La solución analizada está especializada en la inclusión laboral de personas con discapacidad. El proveedor no ha sido identificado con nombre propio en el informe. La auditoría se ha basado en documentación, cuestionario técnico y capturas de pantalla, sin acceso al código fuente ni a los datos reales.

Facilitar la preselección de candidatos con discapacidad ajustados a vacantes específicas, y viceversa, mediante un sistema de recomendación.

El modelo persigue mejorar la eficiencia en los procesos de selección, siempre bajo control humano, asegurando además la coherencia con principios éticos, normativos (RGPD) y sociales de inclusión.

Solución implementada:

Arquitectura:

- Dos modelos independientes entrenados con aprendizaje supervisado:
  - Proceso → Demandante: sugiere personas para vacantes.
  - Demandante → Proceso: sugiere vacantes para personas.

- Datos fuente: base corporativa Integra (≈ 98.000 candidatos y 60.000 vacantes), actualizada cada 24 h y purgada de atributos sensibles (edad, sexo, etc.).
- Variables: matriz de ≈ 1.500 atributos por flujo.
- Salidas: probabilidad de afinidad (0 a 1), ranking, informes individuales, panel de RR. HH. en Power BI, y simuladores tipo "qué pasaría si".
- No se automatizan decisiones, ni existen puntos de corte rígidos; RR. HH. interpreta siempre las recomendaciones.

Resultados obtenidos:

- AUC de hasta 0,97 (traen) y 0,94 (test), pero con posibles síntomas de sobreajuste.
- Alta trazabilidad del proceso mediante registro de versiones, paneles de control, y gestión de incidencias.
- Herramienta útil para identificar brechas de inclusión, sin exclusiones automáticas ni impacto directo en decisiones.
- Muy bien acogido en la organización, con fuerte cultura de gobernanza responsable.

Lecciones aprendidas y áreas de mejora:

Fortalezas:

- Exclusión explícita de variables sensibles.
- Procesamiento técnico sólido, validaciones regulares, y buena capacidad explicativa.
- Participación continua de RR. HH. y diseño de interfaces orientadas a la transparencia y comprensión.
- Formación periódica de los equipos humanos (interpretación de métricas como AUC, DI, SPD, etc.).
- Modelo centrado en la inclusión de colectivos vulnerables, con enfoque de recomendación no determinista.

Áreas de mejora:

1. **Sobreajuste y complejidad:**
   - AUC muy alto con miles de variables indica posible sobreajuste.
   - Recomendación: revisión anual de relevancia de atributos, fusión de redundantes y reducción del número total de variables.
2. **Métricas de equidad ausentes:**
   - No se calculan métricas como Disparate Impact (DI) o Statistical Parity Difference (SPD).
   - Se recomienda incorporarlas en cada ciclo de validación.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

3. **Consentimiento insuficiente:**
   - o Actualmente se basa en disclaimers genéricos. El RGPD exige consentimiento explícito, informado y firmado.
   - o Debe informarse claramente al candidato sobre la finalidad del modelo y su uso.
4. **Evaluación de Impacto:**
   - o No se ha realizado, aunque es obligatoria (art. 35 RGPD) al tratarse de decisiones asistidas que pueden afectar significativamente a personas.
5. **Documentación y trazabilidad:**
   - o Aunque hay paneles y seguimiento, falta sistematización formal del ciclo completo (datos, métricas, decisiones).
   - o Se recomienda conservar versiones con registros de datos, métricas, incidencias y reuniones de validación.
6. **Proxies y sesgos latentes:**
   - o Aunque se eliminan variables sensibles, podrían quedar proxies (p. ej., código postal → nivel socioeconómico).
   - o Se sugiere revisar anualmente correlaciones entre variables clave y atributos sensibles para evitar discriminaciones indirectas.
7. **Preparación para el Reglamento de IA:**
   - o Aunque el modelo fue implantado antes del RIA, deberá adaptarse cuando se produzcan modificaciones sustanciales.
   - o Se recomienda ya aplicar voluntariamente las obligaciones del RIA como marco de referencia ético y normativo.

E008: Gobernanza del dato como fundamento para una IA ética en RR. HH.

Objetivo del análisis: empresa del sector energético que ha desarrollado un modelo integral de Gobernanza del dato con aplicación transversal en la organización, y un foco destacado en el área de Recursos Humanos. La evaluación se ha realizado mediante tres entrevistas en profundidad a perfiles clave del proyecto.

Estudiar cómo la organización ha implementado una arquitectura de Gobernanza del dato sólida y sostenible como base estructural para el uso futuro de IA en procesos laborales. Aunque no se audita un algoritmo específico, se identifican experiencias previas con cribado de CVA y predicción de rotación, dentro de un marco de control humano, ético y regulado.

Enfoque implementado:

Gobernanza organizativa y cultural:

- La gobernanza se integra como una estrategia cultural, no como mera carga técnica.
- Modelo escalonado: madurez progresiva, sin ruptura organizativa.
- Narrativa compartida sobre el valor del dato, que conecta áreas técnicas y de negocio, incluyendo RR. HH.

Arquitectura de roles y órganos:

- Oficina del dato como nodo central de estándares, calidad y trazabilidad.
- Comités de dominio funcional con presencia de perfiles legales, técnicos y de negocio.

• Introducción de data owners (responsables funcionales del dato) y data stewards (garantes técnicos de calidad y trazabilidad).

Gestión del ciclo de vida del dato:

• Énfasis en calidad desde el origen (punto de captura).
• Control riguroso de transformaciones, definiciones validadas y auditoría semántica.
• Aplicación de principios FAIR: Findable (localizable), Accessible, Interoperable, Reusable.

Cultura y accesibilidad del dato:

• Formación continua y transversal en ética, trazabilidad y uso del dato.
• Accesos habilitados por rol, necesidad y registro de uso.
• Iniciativas para empoderar perfiles no técnicos con conocimiento sobre los datos que usan y su impacto.

Casos de uso y aplicaciones tecnológicas:

• RR. HH.: cribado curricular automatizado, predicción de rotación (en fase piloto o ética).
• Operaciones: mantenimiento predictivo, eficiencia energética, trazabilidad.
• Gobernanza del dato: inventario dinámico de algoritmos, identificación de soluciones con IA, formularios de control a proveedores.

Supervisión ética y control responsable:

• Principio rector: “las decisiones sensibles siguen siendo humanas”.
• Comité de IA y Ciberseguridad revisa casos de alto riesgo.
• Cada sistema con IA debe contar con: análisis de riesgos, trazabilidad, responsables asignados.
• Control sobre proveedores externos: formularios obligatorios, cláusulas contractuales, supervisión en cascada.

Retos identificados:

• Resistencia cultural: visión técnica del dato arraigada en algunos equipos.
• Fragmentación semántica: distintas interpretaciones del mismo dato según área.
• Escalabilidad del modelo ético: requiere recursos y coordinación sostenida.
• Opacidad de terceros: dificultad de evaluar algoritmos de proveedores externos.

Conclusiones y aprendizajes:

Fortalezas:

• Modelo realista, transversal y alineado con normas internacionales (ISO/IEC 42001).
• Alta madurez en trazabilidad, calidad, responsabilidad y cultura del dato.
• Gobernanza convertida en activo estratégico, no en freno.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Áreas de mejora:

- Formalizar un código ético específico para IA en RR. HH.
- Automatizar el mantenimiento del inventario de algoritmos.
- Profundizar en auditorías de sesgo e impacto en diversidad.
- Abrir canales de retroalimentación interna sobre decisiones automatizadas.
- Reforzar la supervisión de proveedores mediante herramientas verificables.

Este caso complementa los anteriores al mostrar que, sin un modelo organizativo de gobernanza robusto, ningún sistema de IA en RR. HH. puede ser ético, legal ni eficaz. E008 aporta una visión holística de cómo preparar a una organización para la integración responsable de tecnologías basadas en datos.

E009: Gobernanza del dato: entre interoperabilidad europea y adecuación nacional

Objetivo del análisis: Explorar un modelo de Gobernanza del dato en el ámbito de las políticas activas de empleo. El análisis se basa en una entrevista en profundidad con personal técnico clave de la entidad.

El modelo de Gobernanza del dato propio, sostenible y contextualizado, capaz de:

- Interoperar con marcos europeos como ESCO.
- Responder a necesidades nacionales reales.
- Apoyar procesos como la orientación laboral automatizada.

No se audita un algoritmo en funcionamiento, sino la infraestructura institucional y técnica que sustenta los posibles usos futuros de IA en RR. HH.

Estrategia y solución adoptada:

Gobernanza progresiva y crítica:

- El modelo no parte de un estándar cerrado, sino de una evolución institucional basada en necesidades reales y aprendizajes continuos.
- Se privilegia el desarrollo de soluciones propias, como un catálogo interno de ocupaciones y competencias, ajustado al contexto español y complementario a ESCO.

Interoperabilidad vs adecuación local:

- Se adopta una postura intermedia: ESCO como referencia, pero no como imposición.
- Se identifican importantes brechas semánticas, técnicas y operativas entre los marcos europeos y el sistema español.
- Se crean tablas de equivalencias y mecanismos de integración semántica para asegurar utilidad local sin perder capacidad de interoperar.

Arquitectura técnica y organizativa:

- El catálogo interno alimenta sistemas de orientación laboral automatizada y herramientas de intermediación.
- Se articulan procesos entre técnicos, orientadores y desarrolladores para asegurar usabilidad, coherencia y actualización.
- Se prioriza la sostenibilidad técnica, mediante mantenimiento planificado y arquitectura flexible.

Resultados obtenidos:

- Mejora significativa en la precisión del asesoramiento laboral, al ajustar los datos a realidades locales.
- Capacidad de anticiparse a posibles desajustes provocados por la adopción a ciegas de soluciones externas.
- Fortalecimiento del rol activo de los orientadores en la interpretación y mejora del dato.
- Reconocimiento institucional del dato como un recurso estratégico, no solo administrativo.

Lecciones aprendidas y desafíos superados:

1. **Estandarización sin descontextualización:** El dato debe entenderse en su contexto. ESCO es útil como marco de referencia, pero necesita traducción semántica nacional.
2. **Sostenibilidad técnica:** Las soluciones propias exigen inversión, arquitectura modular y equipos cualificados. El modelo es viable si se planifica a medio plazo.
3. **Coordinación multidisciplinar:** La gobernanza no es solo técnica: requiere puentes entre orientadores, técnicos y responsables institucionales. El lenguaje compartido es clave.
4. **Resistencia frente a imposiciones externas:** Se evita la adopción acrítica de herramientas externas. Se incorpora lo útil, pero se descarta lo que no se alinea con el modelo propio.
5. **Protagonismo de las personas:** El modelo se apoya en la experiencia práctica de los orientadores. Estos detectan vacíos, proponen ajustes y validan categorías.

E009 demuestra que se puede construir una Gobernanza del dato madura, útil y crítica, capaz de:

- Equilibrar estándares internacionales con adecuación nacional.
- Resistir presiones externas mediante diseño propio.
- Involucrar a las personas en el núcleo del sistema, no como usuarios pasivos sino como cocreadores del dato.

Este caso no analiza un modelo de inteligencia artificial concreto, pero ofrece una infraestructura ejemplar sobre la que dichos modelos pueden apoyarse en el futuro de forma ética, responsable y efectiva.

E010: Gobernanza del dato e inteligencia artificial

Objetivo del análisis: Documentar e interpretar la experiencia de E010 en el desarrollo de un modelo

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030 Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

integral de Gobernanza del dato en el ámbito de Recursos Humanos, con especial foco en sus implicaciones éticas, operativas y organizativas.

Enfoque implementado: Modelo progresivo y transversal de Gobernanza del dato, articulado en torno a la Oficina del Dato, comités colegiados y una estrategia basada en casos de uso. Entrevistas presenciales a responsables clave, análisis temático y comparación con marcos normativos de referencia.

Gobernanza organizativa y cultural

La Gobernanza del dato se asume como eje estratégico y catalizador del cambio cultural. Se promueve una pedagogía interna para explicar el valor del dato y vincularlo a resultados organizativos. Se ha trabajado activamente en superar resistencias culturales, integrando a equipos técnicos y funcionales en una narrativa compartida.

Arquitectura de roles y órganos

El modelo se articula en cinco niveles (estratégico, coordinador, funcional, técnico-operativo y usuario final), e incluye:

- Oficina del Dato (núcleo coordinador)
- Comités de Ética, IA y Ciberseguridad
- Data owners (responsables funcionales)
- Data stewards (garantes técnicos de calidad)
- Comités por dominios funcionales

La alta dirección empodera explícitamente la función de gobierno del dato.

Gestión del ciclo de vida del dato

Se trabaja con un enfoque incremental por casos de uso, permitiendo resultados tempranos y sostenibles. Se prioriza la captura en origen con responsabilidad compartida, control de transformaciones y trazabilidad, y una perspectiva orientada a principios FAIR.

Cultura y accesibilidad del dato

Se impulsa una cultura del dato mediante alfabetización interna, talleres éticos y acciones de sensibilización. Se promueve el acceso con responsabilidad, con permisos definidos, trazabilidad de accesos y revisión continua de la sensibilidad del dato.

Casos de uso y aplicaciones tecnológicas:

- **En RR. H.H:** Uso de algoritmos en selección y evaluación, con supervisión humana significativa.
- **En Salud:** Algoritmos para estructurar síntomas desde informes médicos, como base para diagnóstico de enfermedades raras.
- **Gobernanza del dato:** Automatización del inventario de algoritmos, análisis de riesgos y formularios estandarizados para proveedores con IA.

Supervisión ética y control responsable

La ética está institucionalizada. Toda solución con IA pasa por:

- Aprobación del Comité de Ética
- Revisión por Comité de IA y Ciberseguridad
- Documentación sistemática (riesgos, explicabilidad, responsables)
- Contratos con cláusulas específicas para proveedores externos
- Derecho garantizado a la revisión humana

Retos identificados:

- Resistencia cultural a la transversalidad del dato
- Fragmentación semántica entre áreas
- Escalabilidad operativa del modelo ético
- Riesgo por opacidad en soluciones de IA de terceros

Conclusiones y aprendizajes:

- La Gobernanza del dato es una transformación institucional sostenida, no un proyecto puntual.
- La articulación clara de roles y órganos colegiados es clave.
- La ética debe formar parte del diseño, la documentación y la supervisión operativa.
- La cultura y accesibilidad del dato son habilitadores cruciales.
- El modelo de E010 es maduro, replicable y alineado con la norma ISO/IEC 42001, aunque no formalmente certificado.

Fortalezas:

- Liderazgo claro desde la dirección general
- Modelo estructurado y escalable basado en casos de uso
- Alto grado de madurez cultural en torno al dato
- Articulación ética real, operativa y preventiva
- Compromiso con mejora continua y transparencia con terceros

Áreas de mejora:

- Necesidad de formalizar un código ético específico para IA en RR. HH.
- Automatización del inventario de algoritmos.
- Evaluación sistemática del impacto en diversidad e inclusión
- Mecanismos internos de retroalimentación sobre decisiones automatizadas
- Herramientas verificables para auditar soluciones de terceros

A partir del análisis cualitativo de los diez casos de uso y análisis del modelo de Gobernanza del dato, se han identificado una serie de patrones comunes que permiten extraer conocimiento transversal más allá de las particularidades de cada iniciativa.

Con el fin de ofrecer una visión sintética y estructurada del conjunto, a continuación, se presentan

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

varias tablas-resumen que recogen los principales elementos observados en materia de gobernanza del dato, aspectos técnicos, dimensión ética y condiciones organizativas. Estas tablas no atribuyen hechos a ninguna empresa o entidad concreta, agrupan las evidencias más frecuentes o significativas.

Tabla 21: *Elementos observados sobre Gobernanza del dato*

Elemento	Situación observada
Existencia de inventario de algoritmos	En general, ausente o incipiente. Se detecta falta de registros formales y versionado.
Evaluación de Impacto en Protección de Datos	Mayoritariamente no realizada, incluso en sistemas que lo requerirían (art. 35 RGPD).
Trazabilidad del ciclo de vida del dato	Parcial: algunas organizaciones tienen control técnico, pero sin documentación completa.
Segmentación de accesos y roles	Presente en varios casos, pero con limitado control de privilegios y auditoría.
Uso de datos sensibles o proxies	A menudo, excluidos explícitamente, pero no siempre se analizan los proxies.
Reutilización del sistema para otros fines	En algunos casos se observa uso desviado sin evaluación de compatibilidad.
Supervisión de terceros, proveedores externos	Mecanismos informales o no sistemáticos. Falta de cláusulas específicas en contratos.
Formalización de roles y estructuras de gobernanza	En algunos casos se observan definiciones claras de roles (como data owners y stewards) y comités funcionales, lo que facilita la trazabilidad y decisiones colegiadas.
Evaluación ética y supervisión interna	Algunos modelos contemplan la revisión sistemática de algoritmos por comités especializados, con inventarios actualizados y documentación de riesgos.
Cultura del dato y alfabetización	Se identifican iniciativas de formación interna y sensibilización que fortalecen la corresponsabilidad y el uso estratégico del dato.

Fuente: elaboración propia.

Tabla 22: *Aspectos técnicos y metodológicos*

Elemento	Situación observada
Justificación técnica del modelo	Generalmente sólida, con buena elección de algoritmo y validación interna.
Explicabilidad del sistema	Alta en modelos de tipo diagnóstico / bienestar; menor en modelos predictivos.
Inclusión de métricas de equidad	Escasa o nula: no se aplican SPD, DI o similares, salvo en fases piloto.
Formación técnica de usuarios de RR. HH.	Presente en varios casos, pero con brechas en comprensión profunda del modelo.
Simulación o sandbox ético previo	Prácticamente inexistente en fases iniciales.
Documentación de versiones y cambios	Parcial: falta sistematización de decisiones, métricas y procesos de mejora.
Establecimiento de mecanismos de supervisión humana	En algunos casos se implementa el principio de revisión significativa, asegurando que decisiones sensibles no sean totalmente automatizadas.
Uso de formularios de evaluación para proveedores con IA	Se detectan experiencias incipientes con formularios normalizados que exigen detalles sobre funcionamiento del modelo, datos utilizados y análisis de riesgos.
Inventariado técnico de soluciones algorítmicas	Algunos modelos incluyen registros centralizados de algoritmos en uso, con categorización por nivel de riesgo y estado de validación.

Fuente: elaboración propia.

Tabla 23: *Aspectos éticos y organizativos*

Elemento	Situación observada
Consentimiento de los usuarios	En la mayoría de los casos, genérico o implícito. Falta recogida específica.
Comunicación interna sobre el uso del sistema	Variable: alta en proyectos de bienestar; muy baja en usos predictivos.
Participación de las personas afectadas	Escasa: ausencia de mecanismos de feedback o consulta previa.
Supervisión humana en la toma de decisiones	Formalmente presente, aunque no siempre con criterios bien definidos.
Revisión de impacto en colectivos vulnerables	En general, no realizada o sistematizada.
Marco ético o código de conducta específico	Ausente o en elaboración. Algunas organizaciones lo suplen con buena práctica informal.
Aprobación previa por comités especializados	En algunos casos, se condiciona el despliegue de soluciones con IA a la aprobación explícita por comités de ética, IA o ciberseguridad.
Cláusulas éticas en contratos con terceros	Se detectan buenas prácticas emergentes que incluyen obligaciones éticas y de explicabilidad para proveedores tecnológicos.
Canales de retroalimentación interna	Ausencia generalizada, aunque empiezan a explorarse mecanismos que permitan a empleados reportar preocupaciones o errores derivados del uso automatizado de datos.

Fuente: elaboración propia.

Tabla 24: *Factores organizativos que facilitan o dificultan el uso ético de IA en RR. HH.*

Factor organizativo	Impacto observado
Existencia de cultura del dato	Las organizaciones con narrativa común sobre el valor del dato muestran más trazabilidad y control.
Implicación transversal de áreas (RR. HH., IT, Legal…)	Mejora la calidad del diseño y reduce riesgos éticos. Falta de coordinación genera puntos ciegos.
Participación de la plantilla en el diseño o prueba	Casi nula. Las iniciativas sin diálogo con las personas afectadas tienden a menor aceptación interna.
Presencia de estructuras éticas o comités internos	Ausencia generalizada. Su creación correlaciona con mayor capacidad para evaluar riesgos futuros.
Madurez documental y procedimental	Las organizaciones con estándares de calidad y procesos definidos escapan del "efecto piloto perpetuo".
Dependencia del proveedor externo	Alta dependencia sin herramientas de auditoría reduce la soberanía técnica y jurídica de la organización.
Formalización de roles específicos para la gobernanza del dato	La existencia de perfiles como data owners y stewards mejora la trazabilidad, la rendición de cuentas y la consistencia técnica.
Adopción de enfoque incremental por casos de uso	Favorece la escalabilidad y legitimidad del modelo, al generar resultados rápidos y facilitar aprendizaje organizativo progresivo.
Integración de principios éticos desde fases tempranas	Las organizaciones que consideran la ética desde el diseño técnico, muestran mayor coherencia y menos conflictos en fases de implementación.

Fuente: elaboración propia.

Las tablas anteriores permiten reconocer tanto los avances como las fragilidades que están presentes hoy en el despliegue de sistemas de IA en el ámbito de Recursos Humanos.

Frente a la heterogeneidad de enfoques, se detectan desafíos compartidos que requieren respuestas

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

estructurales: trazabilidad del dato, transparencia organizativa, equidad algorítmica, y control humano efectivo, entre otros.

Sobre esta base empírica y sin ánimo de prescripción única, el siguiente apartado formula una serie de conclusiones generales, reflexiones estratégicas y recomendaciones prácticas, que buscan orientar un desarrollo más ético, legal y útil de la IA en la gestión de personas.

3.3 Conclusiones empíricas

3.3.1 Lecciones aprendidas: fundamentos para una IA en RR. HH. ética y efectiva

El análisis transversal de los casos presentados ha permitido detectar tanto prácticas consolidadas como fragilidades comunes en el uso de IA en RR. HH. Sobre esta base empírica, a continuación, se formulan una serie de conclusiones estructuradas en torno a tres ejes: aprendizajes clave, mapa de riesgos y oportunidades, y recomendaciones aplicables.

- **Madurez organizativa previa** (especialmente en Gobernanza del dato): es un factor clave para el éxito y la sostenibilidad de cualquier sistema basado en IA.
- **Explicabilidad y supervisión humana:** no son atributos opcionales, sino condiciones imprescindibles para asegurar confianza y legitimidad.
- **Sensibilidad ética creciente:** muchas organizaciones han mostrado este nuevo tipo de sensibilidad algorítmica, aunque no siempre va acompañada de los procedimientos formales exigidos (consentimiento, auditorías, trazabilidad).
- **Exclusión de variables sensibles** (género, edad, origen, etc.): no garantiza por sí sola la equidad algorítmica y es necesario considerar también sus proxies y efectos indirectos.

Las mejores prácticas observadas no derivan de la complejidad técnica del modelo, sino de la claridad en sus objetivos, la transparencia de su uso y la responsabilidad con que se gestiona su impacto.

Estas lecciones no prescriben, pero permiten comprender qué factores favorecen el despliegue responsable de IA en RR. HH. y sirven de base para las recomendaciones normativas y prácticas que se exponen a continuación.

3.3.2 Mapa de riesgos y oportunidades: claves para anticipar el futuro

A continuación, se sintetizan los principales desafíos y potenciales positivos identificados a lo largo de los análisis:

Riesgos recurrentes

- Reutilización de modelos o datos para fines no previstos, sin evaluación previa.
- Falta de documentación estructurada, especialmente en modelos en fase piloto o sin responsable asignado.
- Procesos de consentimiento difusos, basados en cláusulas genéricas.
- Débil uso de métricas de equidad, lo que dificulta identificar sesgos emergentes.
- Exposición injusta de figuras intermedias (p. ej. mánagers), si los datos no consideran factores estructurales.
- Dependencia de proveedores sin mecanismos de control o auditoría verificable.

Oportunidades clave

• Diseñar la IA como un apoyo estratégico, no como un mecanismo de automatización opaca.
• Fortalecer la cultura del dato ético desde los equipos de RR. HH. hacia toda la organización.
• Crear inventarios de algoritmos y comités de evaluación permanente, más allá del cumplimiento normativo.
• Promover el desarrollo conjunto de soluciones con los colectivos afectados, garantizando participación activa.
• Transformar la Gobernanza del dato en un activo organizativo: trazable, auditable y comprensible para todos los perfiles.
• Aprovechar la IA para visibilizar desigualdades ocultas y orientar políticas de equidad y diversidad.

3.3.3 Recomendaciones prácticas para desplegar la IA en RR. HH. con garantías

A la luz del análisis empírico, se proponen las siguientes orientaciones para implementar sistemas de IA en el ámbito de los Recursos Humanos de forma ética, eficaz y legalmente conforme:

1. Establecer un marco de Gobernanza del dato y algoritmos, con responsables definidos, criterios de acceso, trazabilidad, y participación multidisciplinar (RR. HH., legal, IT y ética).

2. Realizar siempre una Evaluación de Impacto en Protección de Datos cuando el sistema afecte a personas en decisiones significativas (selección, promoción, despido).

3. Recoger consentimiento explícito, informado y revocable, especialmente en usos individuales o automatizados. Evitar los disclaimers genéricos.

4. Documentar todo el ciclo de vida del sistema: fuentes de datos, lógica del modelo, versiones, métricas de validación, decisiones y responsables.

5. Incorporar métricas de equidad algorítmica (DI, SPD, Predictive Parity, Equalized Odds), y analizarlas desagregadas por colectivos vulnerables.

6. Formar al equipo de RR. HH. y gestores en interpretación responsable de resultados, límites del modelo y comunicación interna.

7. Garantizar el derecho a explicación, revisión humana y apelación en todo uso que afecte a decisiones personales.

8. Respetar el principio de limitación de la finalidad (art. 5.1.b del RGPD): si se reutiliza un algoritmo o conjunto de datos para un fin distinto al previsto originalmente, es obligatorio justificar la compatibilidad del nuevo fin, documentar el cambio, y recabar un nuevo consentimiento o base legal adecuada. La falta de transparencia o re uso indebido puede constituir una infracción sancionable.

La inteligencia artificial aplicada en Recursos Humanos solo podrá considerarse legítima si se inserta en una Gobernanza del dato robusta, una cultura organizativa transparente y un marco ético que sitúe a las personas en el centro de la decisión.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

4. RECOMENDACIONES PARA LA CERTIFICACIÓN DE SISTEMAS DE IA APLICADOS A RR. HH. [9]

En la actualidad, el desarrollo y despliegue de sistemas de inteligencia artificial en el ámbito de Recursos Humanos está avanzando a gran velocidad, transformando profundamente procesos como la selección, la formación, la evaluación del desempeño, la planificación de carreras o incluso la salud en el trabajo, incluyendo aspectos como la ergonomía o la prevención de riesgos laborales.

4.1 Introducción y contexto

Este despliegue abre nuevas oportunidades para la eficiencia y la innovación, pero también introduce retos significativos en materia de transparencia, equidad, protección de datos y toma de decisiones automatizadas, que inciden directamente sobre los derechos, trayectorias y expectativas profesionales de las personas trabajadoras. En consecuencia, garantizar la responsabilidad, trazabilidad y confiabilidad del uso de la IA en los procesos de RR. HH. se ha convertido en una prioridad para cualquier organización comprometida con un futuro laboral justo, inclusivo y sostenible.

Las organizaciones deben no solo cumplir con los marcos normativos vigentes —como el RIA o la norma ISO/IEC 42001:2023, que establece sistemas de gestión para la IA—, sino también generar confianza entre empleados, candidatos y otros grupos de interés. Esta confianza se construye desde el cumplimiento legal y desde la ejemplaridad ética y la capacidad de demostrar un uso responsable, explicable y auditado de las tecnologías. Las directrices éticas europeas para una IA confiable, junto con normas como la ISO/IEC 23894:2023 (gestión de riesgos en IA) o la ISO/IEC TR 24368:2022 (consideraciones éticas), ofrecen un marco de referencia útil para avanzar en esta dirección.

Frente a estos desafíos, se propone avanzar hacia la creación de un sistema de certificación específico para la IA aplicada a los recursos humanos. El objetivo no es únicamente formular un conjunto de recomendaciones técnicas, sino diseñar un marco evolutivo y escalable que permita a las organizaciones progresar por etapas, reconociendo sus avances mediante un modelo de certificación estructurado en niveles de madurez ética y técnica.

Esta propuesta se fundamenta en dos pilares complementarios:

1. El diseño de un **sello ético y de calidad**, que reconozca públicamente los esfuerzos en transparencia, gobernanza algorítmica y respeto a los derechos fundamentales en entornos laborales.

2. La definición de los **elementos clave del sistema de certificación algorítmica**, incluyendo

[9] **Nota**: Este capítulo, así como su correspondiente anexo, se ha elaborado gracias a la colaboración de Guillermo Butler, CEO de ComplianceCMS, y las aportaciones de Manuel Zaera, COO & Head of HC y Elena Martín de Diego, AI Certification Project Lead, ambos de Zertia.

requisitos sobre los datos, los modelos, el impacto organizativo, la participación de terceros y la accesibilidad del sistema para las personas usuarias.

Asimismo, el enfoque gradual propuesto —desde un compromiso inicial hasta niveles avanzados de liderazgo— busca evitar barreras de entrada, incentivar la mejora continua y facilitar la incorporación de entidades de todos los tamaños y sectores. Esta lógica de progresión, inspirada en modelos de madurez organizativa, se alinea tanto con las recomendaciones del NIST AI Risk Management Framework como con los principios definidos por los Objetivos de Desarrollo Sostenible (especialmente los ODS 8, 9 y 10, sobre trabajo decente, innovación e igualdad).

En definitiva, este capítulo ofrece una hoja de ruta clara, técnica y ética para aquellas organizaciones que desean construir un ecosistema laboral más inclusivo, confiable y responsable. La certificación no es un fin en sí mismo: es una herramienta para fomentar una cultura organizativa donde la inteligencia artificial sirva al desarrollo humano, al empoderamiento de las personas y a la consolidación de estándares éticos compartidos.

4.2 Fundamentos para un sello de calidad “ética” en IA

La creciente integración de sistemas de inteligencia artificial en la gestión de personas transforma de forma acelerada el entorno laboral, automatizando procesos que tradicionalmente dependían del juicio humano —como la selección, la promoción o la evaluación del desempeño—. Este cambio requiere una reflexión profunda sobre los principios que deben guiar el uso responsable y justo de estas tecnologías.

En este contexto, un sello ético de calidad se plantea como una herramienta clave para distinguir a las organizaciones comprometidas con una IA centrada en las personas, que respete la equidad, los derechos fundamentales y la transparencia algorítmica.

Lejos de limitarse a un emblema reputacional, este sello debe tener una función transformadora: reconocer públicamente el compromiso ético real, impulsar una cultura organizativa basada en la supervisión crítica y promover una adopción progresiva de buenas prácticas en IA.

4.2.1 Principios rectores

El sello se fundamenta en un conjunto de principios ampliamente reconocidos en marcos internacionales de ética digital:

- **Justicia y no discriminación,** para prevenir impactos diferenciales injustificados.
- **Transparencia y explicabilidad,** que permitan comprender las decisiones automatizadas.
- **Supervisión humana significativa,** como garantía de control e intervención.
- **Responsabilidad y rendición de cuentas,** para asegurar trazabilidad y respuesta ante errores o sesgos.

Estos principios están alineados con directrices como las de la Comisión Europea para una IA confiable (2019), el enfoque del IEEE Ethically Aligned Design, y los fundamentos recogidos en la ISO/IEC TR 24368:2022, la ISO/IEC 22989:2023 y el reciente Convenio sobre Inteligencia Artificial del Consejo de Europa (2024).

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

4.2.2 Adaptación al contexto laboral

El sello debe incorporar una adaptación específica al ámbito de Recursos Humanos, reconociendo cómo las decisiones algorítmicas que se toman en este entorno tienen consecuencias inmediatas, personales y a menudo irreversibles.

Por ello, se requiere:

- Sensibilidad ante las dinámicas de poder y vulnerabilidad laboral.
- Participación activa de los colectivos afectados.
- Capacidad de impugnación efectiva de decisiones automatizadas.
- Revisión ética que considere la equidad interna en cada proceso (selección, formación, promoción…).

Esta especificidad hace que el sello no sea una simple traslación de modelos generales: se trata de una construcción contextualizada a la realidad organizativa y relacional del trabajo.

4.2.3 Inspiración en marcos existentes

Para asegurar su credibilidad, el desarrollo del sello debe alinearse con normas internacionales consolidadas y mecanismos de gobernanza ya existentes. Entre ellas:

- **ISO/IEC 42001:2023-** Sistemas de gestión para IA.
- **ISO/IEC 23894:2023-** Gestión de riesgos en IA.
- **TR 24027:2023-** Evaluación de sesgos algorítmicos.
- **RIA-** Modelos regulatorios como el europeo, que identifica los sistemas de IA en RR. HH. como de alto riesgo.

Este alineamiento técnico-normativo facilitará la interoperabilidad del sello con auditorías externas, requisitos legales o procesos de verificación ya presentes en las organizaciones.

4.2.4 Dimensión organizativa y transformadora

Más allá de la certificación como acto final, el sello debe entenderse como una herramienta de transformación cultural y organizativa. Su aplicación debe:

- Fomentar la **alfabetización algorítmica** de los equipos.
- Facilitar la creación de **comités éticos internos o grupos multidisciplinares** (con perfiles técnicos, legales, de RR.HH. y sociológicos).
- Promover procesos de deliberación ética y revisión periódica del impacto de la IA.

En línea con el NIST AI Risk Management Framework, se recomienda acompañar el despliegue del sello con estrategias de sensibilización, formación y documentación adaptadas al lenguaje de los distintos actores implicados.

4.2.5 Más allá de la certificación

Finalmente, los fundamentos del sello no deben limitarse a la verificación de requisitos formales, sino

reflejar una visión humanista de la tecnología, orientada al fortalecimiento de los derechos laborales, la justicia organizativa y el progreso social.

La inteligencia artificial en el ámbito de la gestión de personas debe concebirse no como un sustituto de la ética profesional, sino como un ámbito donde esta se proyecta con mayor necesidad.

4.3 Arquitectura del sello de calidad por niveles

4.3.1 Enfoque progresivo para fomentar la mejora continua

Uno de los principales retos al establecer un sistema de certificación en inteligencia artificial aplicada a la gestión de personas es encontrar un equilibrio adecuado entre exigencia ética y viabilidad operativa. Un modelo excesivamente rígido podría disuadir a las organizaciones de participar, mientras que uno demasiado laxo correría el riesgo de banalizar el compromiso ético. Para afrontar este dilema, se propone una arquitectura escalonada de certificación basada en niveles de madurez, que permita a cada organización avanzar progresivamente según su grado de alineamiento con los principios éticos y técnicos establecidos.

Este enfoque, inspirado en modelos como el AI Risk Management Framework del NIST y el AI Maturity Model de Gartner, permite reconocer avances tangibles y fomenta una lógica de mejora continua. Además, refuerza la idea de que la certificación no es un acto puntual, sino un proceso evolutivo que refleja la madurez institucional en el uso responsable de la IA.

4.3.2 Descripción de los niveles

Nivel Bronce: Compromiso inicial

Reconoce a las organizaciones que expresan su compromiso con una IA ética en RR. HH. Para este nivel se requiere:

- Una autoevaluación estructurada de riesgos éticos y funcionales.
- Un plan de acción definido y aprobado por la dirección.
- La adopción pública de principios rectores mediante una Política de Cumplimiento Ético.
- Adhesión formal al modelo de certificación.

Este nivel permite visibilizar una voluntad clara de alineamiento, aún sin implementación completa.

Nivel Plata: Implementación activa y primer compromiso estructural

Acredita que la organización ha comenzado a implementar de forma activa su plan de mejora:

- Evidencia documental de auditorías internas o externas.
- Protocolos de gobernanza en uso y revisión del sistema de IA.
- Formación inicial de los equipos implicados.
- Procedimientos internos para la revisión y mitigación de sesgos.

El enfoque en esta etapa es la maduración operativa del compromiso ético inicial.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Nivel Oro: Excelencia ética

Este nivel reconoce a organizaciones con un alto grado de cumplimiento normativo y ético, que han consolidado un sistema trazable, transparente y supervisado. Se requiere:

- Cumplimiento técnico completo de los estándares aplicables.
- Canales efectivos para el ejercicio de derechos digitales y apelación.
- Auditorías periódicas por entidades independientes.
- Participación activa de los grupos afectados (empleados, sindicatos, etc.).
- Revisión ética integrada en el diseño, desarrollo y uso del sistema.

El nivel Oro constituye el estándar de referencia para entornos exigentes y altamente regulados.

Nivel Platino: Liderazgo (opcional)

Diseñado para organizaciones que no solo cumplen, sino que lideran activamente en el ámbito de la ética algorítmica. Se espera que:

- Participen en redes de innovación y en la elaboración conjunta de normas.
- Publiquen informes de transparencia y compartan buenas prácticas.
- Contribuyan a la elaboración de estándares sectoriales o intersectoriales.
- Evalúen de forma proactiva el impacto social y laboral de sus sistemas.

Este nivel refuerza el principio de corresponsabilidad sectorial, donde las entidades certificadas ayudan a elevar los estándares comunes [10].

4.3.3 Modularidad y enfoque por procesos

El sistema de certificación puede diseñarse con modularidad, permitiendo la evaluación diferenciada por áreas funcionales (por ejemplo: selección, formación, evaluación del desempeño). Este modularidad:

- Facilita la entrada progresiva al sistema, especialmente para pymes o sectores específicos.
- Permite certificar partes del sistema sin necesidad de alcanzar una cobertura total inicial.
- Favorece la visibilidad de mejoras continuas y escalables dentro de las organizaciones[19].

4.3.4 Equivalencias con marcos de madurez organizacional

Para reforzar su solidez, la arquitectura propuesta puede alinearse con modelos de madurez organizacional, como los utilizados en transformación digital (el Digital Acceleration Index de BCG) o en sistemas de gestión (el definido por la norma ISO/IEC 33001). Esto facilitaría su integración con auditorías previas, evitaría duplicidades y permitiría establecer sinergias con otros sistemas de gestión, especialmente si la organización ya trabaja bajo esquemas como el ISO/IEC 42001.

---

[10] **Nota aclaratoria sobre los niveles y la certificación:** la obtención de los niveles no implica automáticamente una certificación formal completa según criterios legales, pero sí permite: 1. Emitir declaraciones de conformidad proporcionadas al nivel alcanzado; 2. Validarlas mediante procesos verificables e independientes y 3. Usarlas como base para transitar hacia una certificación integral, especialmente en niveles Oro y Platino.

4.4 Elementos clave del sistema de certificación algorítmica

4.4.1 Una mirada integral al ciclo de vida del sistema algorítmico

Un sistema de certificación robusto y ético de inteligencia artificial debe ofrecer una visión integral del ciclo de vida de los sistemas algorítmicos, considerando no solo sus componentes técnicos, sino también sus contextos de uso, impactos organizativos y condiciones de supervisión y gobernanza. Esta perspectiva holística se alinea con marcos como la ISO/IEC 23894:2023 y la ISO/IEC 42001:2023, que destacan la necesidad de evaluar los riesgos y responsabilidades en cada fase del sistema.

A continuación, se detallan los principales bloques que conforman los elementos clave del sistema de certificación, orientados a garantizar un uso justo, transparente, robusto y socialmente responsable de la IA en los procesos laborales.

4.4.2 Evaluación de los datos: calidad, diversidad y gobernanza

Los datos son el punto de partida de cualquier sistema de IA. Su origen, estructura y uso determinan en gran medida los resultados y sesgos del sistema.

Un sistema de certificación tendría que evaluar:

- La licitud, relevancia y actualización de los datos utilizados.
- La existencia de medidas de **protección de datos** (como la anonimización, minimización, conservación limitada).
- La **representatividad de los datos** según variables relevantes (género, edad, etnia u otros) para evitar sesgos estructurales, en línea con la ISO/IEC TR 24027:2023.
- La **trazabilidad** de los datos a lo largo del tiempo.
- La existencia de una **Gobernanza del dato formalizada**, integrada en el sistema de gestión organizacional.

4.4.3 Evaluación del modelo: transparencia, sesgos y validación

El modelo es el componente técnico central del sistema de IA, y debe ser evaluado tanto en su diseño como en su comportamiento en uso real.

La certificación del sistema de inteligencia artificial verificará:

- La **validación técnica** en contextos reales y la documentación de indicadores clave (precisión, robustez, generalización).
- La existencia de **mecanismos de explicabilidad técnica y organizativa**, ajustados a distintos perfiles de usuarios.
- La ejecución de pruebas para la **detección de sesgos algorítmicos**, tanto estadísticas como empíricas.
- La existencia de protocolos de revisión periódica y control de desviaciones.
- La trazabilidad del modelo en todo su ciclo de vida, incluyendo versiones, actualizaciones y responsables de cambios.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Se valorará la diferencia entre modelos desarrollados internamente y aquellos adquiridos a terceros, exigiendo documentación técnica en ambos casos.

4.4.4 Evaluación del impacto organizativo y de gobernanza

El funcionamiento ético de un sistema algorítmico no depende solo de su diseño técnico, sino también del contexto organizativo en el que opera. Por eso, se revisará:

- La existencia de **canales internos de supervisión y rendición de cuentas**.
- La formación del personal en el uso, interpretación y revisión del sistema.
- La trazabilidad de las decisiones tomadas con intervención algorítmica.
- La posibilidad de **impugnación o apelación por parte de las personas afectadas,** (Revisar el apartado referido al Marco Legal).
- La inclusión de cláusulas éticas en contratos con terceros proveedores.

Una gobernanza sólida debe promover la deliberación interna, el control humano significativo y una cultura organizativa proactiva en materia de ética tecnológica.

4.4.5 Participación de terceros: independencia y supervisión

Un componente esencial para dotar de legitimidad al sistema de certificación es la implicación de actores externos. Esto puede adoptar distintas formas:

- Realización de auditorías por **entidades independientes certificadas.**
- Creación o consulta a **comités éticos internos o interdisciplinares.**
- Participación de **agentes sociales** u organizaciones representativas de los trabajadores en la supervisión del impacto.
- Publicación de **informes de transparencia**, como sugiere el OECD Framework for the Classification of AI Systems, que promueve la supervisión proporcional según el nivel de riesgo y exposición.

La interacción con terceros introduce una dimensión de vigilancia ética externa que refuerza la credibilidad del sello de calidad y amplía el control más allá del ámbito corporativo.

4.4.6 Incorporación de criterios de usabilidad y accesibilidad del sistema para empleados

Un elemento frecuentemente subestimado en los procesos de certificación algorítmica es el grado en que las herramientas de IA son comprensibles y utilizables por quienes interactúan con ellas en el entorno laboral, incluyendo perfiles no técnicos.

La usabilidad y la accesibilidad no solo son atributos deseables desde una perspectiva técnica o de experiencia de usuario, sino que constituyen garantías éticas fundamentales para que los sistemas no generen exclusión, incomprensión o dependencia tecnológica.

Desde el punto de vista de la certificación, se necesitan criterios que aseguren que:

- Las interfaces utilizadas por empleados y gestores son **intuitivas y proporcionales al nivel de formación técnica disponible.**

• Se ofrecen **explicaciones comprensibles** sobre los resultados del sistema, adaptadas al lenguaje y roles del entorno organizativo.
• Se contemplan **ajustes razonables y configuraciones accesibles** para personas con diversidad funcional, siguiendo estándares de accesibilidad digital (por ejemplo, WCAG) y normas de usabilidad reconocidas como la ISO/IEC 25066:2016, que orienta sobre cómo evaluar y documentar la experiencia de uso en sistemas interactivos.
• Se evalúa de forma empírica la **experiencia de los usuarios** a través de pruebas de interacción o encuestas de satisfacción.
• Se documenta si la herramienta permite que los empleados **comprendan cuándo interviene un sistema algorítmico** y cuál es su rol en la decisión final.

Cuando los sistemas de IA pueden afectar directa o indirectamente a trayectorias profesionales, la transparencia no puede limitarse a los desarrolladores o a los responsables de TI. La certificación debe garantizar que las personas destinatarias de los resultados algorítmicos puedan ejercer agencia informada sobre su interpretación, respuesta o impugnación. De este modo, la usabilidad y la accesibilidad no se entienden como aspectos operativos secundarios, sino como condiciones estructurales de equidad y empoderamiento en el entorno de trabajo.

4.5 Recomendaciones para la implantación progresiva

4.5.1 Implantar la certificación como proceso organizativo, no como formalismo

La certificación ética y técnica en inteligencia artificial aplicada a los recursos humanos no debe abordarse como una tarea puntual o una exigencia externa. Para que tenga impacto real y sostenido, es necesario que se entienda como un proceso organizativo gradual, integrado en la evolución estratégica de cada entidad. Esta perspectiva es coherente con normas como la ISO/IEC 42001:2023 (gestión de sistemas de IA), que establece principios de gestión progresiva, basada en ciclos de mejora continua, y con los principios de gobernanza flexible propuestos por la OECD y el NIST AI Risk Management Framework.

Estas recomendaciones pretenden ofrecer una hoja de ruta adaptable para que diferentes organizaciones, independientemente de su tamaño o sector, puedan avanzar de forma estructurada hacia la obtención del sello propuesto.

La certificación ética y técnica en inteligencia artificial aplicada a los recursos humanos no debe abordarse como una tarea puntual o una exigencia externa. Para que tenga impacto real y sostenido, es necesario que se entienda como un proceso organizativo gradual, no como un formalismo, integrado en la evolución estratégica de cada entidad.

Esta perspectiva es coherente con normas como la ISO/IEC 42001:2023 (gestión de sistemas de IA), que establece principios de gestión progresiva, basada en ciclos de mejora continua, y con los principios de gobernanza flexible propuestos por la OECD y el NIST AI Risk Management Framework.

Estas recomendaciones pretenden ofrecer una hoja de ruta adaptable para que diferentes organizaciones, independientemente de su tamaño o sector, puedan avanzar de forma estructurada hacia la obtención del sello propuesto.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

4.5.2 Fases de implantación: de la reflexión al reconocimiento

Se propone una hoja de ruta en cuatro fases, adaptables a distintos tipos de organización:

1. Diagnóstico ético-técnico inicial

- Identificar los sistemas de IA ya implantados en los procesos de RR. HH.
- Analizar su nivel de automatización, sensibilidad ética y madurez técnica.
- Evaluar la Gobernanza del dato, el ciclo de vida algorítmico y el grado de participación de los equipos implicados.

2. Alineamiento estratégico y definición de objetivos

- Conectar el uso de la IA con los valores corporativos y los compromisos éticos existentes.
- Establecer objetivos específicos a corto, medio y largo plazo, integrados en el plan estratégico de RR. HH. o digitalización.
- Evaluar el impacto sobre la cultura organizativa y los modelos de decisión interna.

3. Implementación del plan de acción y medición de avances

- Ejecutar auditorías internas, formaciones específicas, actualizaciones de protocolos y mecanismos de explicabilidad.
- Documentar evidencias y mejoras introducidas.
- Medir el progreso con indicadores cualitativos y cuantitativos, tales como: nivel de cumplimiento, percepción del personal, capacidad de revisión, etc.

4. Revisión externa y certificación del nivel alcanzado

- Iniciar el proceso formal de auditoría externa según el nivel (Bronce, Plata, Oro o Platino).
- Realizar entrevistas, revisión documental, pruebas funcionales y verificación de evidencias.
- Emitir declaración de conformidad y registrar las observaciones para el siguiente ciclo de mejora.

4.5.3 Adaptar las recomendaciones al tipo de organización

Las características estructurales, culturales y tecnológicas de cada entidad influirán en cómo se lleva a cabo este proceso. Por ello, se ofrecen recomendaciones específicas según el tipo de organización:

• Grandes empresas

o Recomendable abordar la certificación desde una perspectiva transversal, implicando múltiples departamentos (recursos humanos., TI, compliance, ética).
o Se sugiere crear comités internos de ética digital, designar un responsable de inteligencia artificial en recursos humanos y publicar informes de avance periódicos.

• **Pequeña y mediana empresa (pymes)**

o Pueden iniciar con una certificación modular por proceso (por ejemplo, selección) y apoyarse en herramientas estandarizadas.
o Se recomienda acudir a redes de mentorización o servicios de acompañamiento externo.

• **Sector Público y administraciones**

o La certificación puede alinearse con compromisos de transparencia, derechos digitales y eficiencia.
o Es aconsejable integrarla con marcos de gobernanza ya existentes (por ejemplo, ISO/IEC 27014:2020) y favorecer la colaboración interinstitucional.

4.5.4 Herramientas y recursos de apoyo

El despliegue progresivo puede apoyarse en materiales y recursos complementarios, como:

• Guías de autoevaluación ética.
• Plantillas de documentación para auditorías y validaciones.
• Manuales de buenas prácticas adaptados al contexto de RR. HH.
• Ejemplos reales o simulados de implementación por niveles.
• Plataformas digitales para seguimiento del avance y autodiagnóstico, como las que propone la OECD en su catálogo de herramientas.

Estos materiales deberán actualizarse regularmente y adaptarse a nuevas normativas, avances tecnológicos o aprendizajes derivados de su aplicación en distintos entornos.

4.5.5 Acompañamiento y/o mentorización para avanzar de un nivel a otro

Uno de los factores críticos para garantizar que la certificación ética en IA no se convierta en una herramienta excluyente es ofrecer estructuras de acompañamiento adaptadas al punto de partida de cada organización.

Esta lógica de progresión escalonada puede enriquecerse mediante mecanismos de mentorización y asesoramiento externo, orientados a facilitar el tránsito entre niveles y a superar los retos específicos que surgen en cada etapa del proceso.

Este acompañamiento puede adoptar diversas formas:

• **Redes de organizaciones mentoras,** donde empresas u organismos con mayor madurez comparten experiencias, aprendizajes y herramientas con entidades que inician el proceso.

• **Servicios de apoyo técnico o ético** prestados por agentes públicos, universidades, asociaciones profesionales o centros de innovación especializados.

• **Modelos de certificación guiada,** en los que el proceso no solo concluye con una verificación, sino que incluye sesiones intermedias de orientación, revisión de avances y recomendaciones personalizadas.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030 Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

• **Espacios compartidos de aprendizaje,** como comunidades de práctica, laboratorios éticos o grupos sectoriales, donde se puedan discutir buenas prácticas y obstáculos comunes.

El acompañamiento no implica suavizar los requisitos, sino hacerlos alcanzables sin comprometer la exigencia ética. Además, refuerza el carácter colaborativo del sello, al fomentar una cultura de mejora continua, solidaridad Inter organizativa y rendición de cuentas compartida.

Este tipo de dinámicas ya han demostrado su eficacia en otros entornos de certificación (como calidad, sostenibilidad o digitalización), y pueden ser clave para democratizar el acceso a la certificación ética en IA aplicada a RR. HH. Iniciativas como los Digital Innovation Hubs promovidos por el Programa Europa Digital de la Comisión Europea ofrecen un marco consolidado para este tipo de acompañamiento estructurado, combinando asesoramiento, formación y orientación ética y tecnológica.

4.6 Mecanismos de verificación y seguimiento

4.6.1 Más allá del sello: garantizar confianza a lo largo del tiempo

Una certificación ética en inteligencia artificial solo adquiere valor real cuando cuenta con mecanismos sólidos de verificación y seguimiento, que legitimen su otorgamiento inicial y aseguren su vigencia, adaptabilidad y coherencia a lo largo del tiempo. En un entorno tecnológico dinámico como el de la IA, donde los modelos evolucionan y los contextos cambian, los sistemas de verificación deben combinar rigor técnico con agilidad operativa. Normas como la ISO/IEC 42001:2023 y marcos como el NIST AI Risk Management Framework destacan la importancia de integrar procesos de revisión continua, interna y externa, en los sistemas de gestión de IA.

El sello de calidad propuesto no debe entenderse como un reconocimiento estático, sino como parte de un proceso constante de mejora, revisión y rendición de cuentas.

4.6.2 Verificación interna: autoevaluaciones sistemáticas

El primer nivel de control se basa en autoevaluaciones ético-técnicas periódicas, realizadas por la propia organización. Estas evaluaciones permiten:

- **Monitorear el grado de cumplimiento** respecto a los principios y nivel de certificación alcanzado.
- **Detectar desviaciones**, nuevas áreas de riesgo o transformaciones relevantes en el sistema de inteligencia artificial.
- Activar alertas internas que deriven en **revisiones o acciones correctivas.**

La ISO 19011:2018, que proporciona directrices para la auditoría de sistemas de gestión, ofrece una base sólida para estructurar estas autoevaluaciones, garantizando que sean completas, trazables y reproducibles.

4.6.3 Verificación externa: auditorías independientes

La verificación externa independiente es el segundo pilar del sistema. Las auditorías pueden ser realizadas por entidades certificadoras acreditadas o agencias reguladoras según estándares

reconocidos. Estas auditorías deben ser proporcionales al nivel de certificación (Bronce, Plata, Oro, Platino) y contemplar:

- Evaluación documental técnica y organizativa.
- Entrevistas con responsables y usuarios de los sistemas.
- Pruebas de trazabilidad y validación de decisiones algorítmicas.
- Validación de criterios de equidad, transparencia y explicabilidad.

Herramientas como la AI Audit Tool, desarrollada por AlgorithmWatch, o directrices más técnicas como las de la ISO/IEC 27007:2020, ofrecen criterios útiles para realizar auditorías que combinen revisión técnica y análisis desde la perspectiva de derechos fundamentales.

La frecuencia y profundidad de las auditorías podrá variar en función del nivel del sello alcanzado, del impacto del sistema y de los cambios organizativos o tecnológicos introducidos.

4.6.4 Declaraciones de conformidad y sellos temporales

Para garantizar la validez dinámica de la certificación, se propone que el sello:

- Tenga una vigencia limitada (2 años para Bronce y Plata, 3 años para Oro y Platino).
- Incorpore una revisión intermedia simplificada a mitad de ciclo.
- Permita la revocación o suspensión del sello en caso de incumplimientos graves o cambios no documentados.

Este modelo evita que el sello se convierta en un elemento simbólico estático, y fomenta un enfoque de vigilancia activa y mejora continua, en línea con las recomendaciones del OECD AI Framework.

4.6.5 Sistemas de seguimiento digital y trazabilidad

Para facilitar la gestión del proceso, se recomienda el desarrollo o adopción de plataformas digitales de seguimiento, que permitan:

- Registrar el estado de avance hacia la certificación.
- Documentar evidencias y acciones correctoras.
- Visualizar el historial de auditorías y renovaciones.
- Integrarse con sistemas de gestión documental, compliance y control de calidad.

Estas plataformas pueden integrarse con dashboards internos o repositorios documentales, y en el caso de administraciones públicas o grandes organizaciones, puede considerarse una apertura parcial de estos registros, como forma de rendición de cuentas.

4.6.6 Mecanismos de actualización y revisión del sistema

El último componente esencial del sistema de seguimiento es la revisión periódica del modelo de certificación en sí mismo. Se propone:

- Revisión anual o bianual de criterios y requisitos del sello.
- Nuevos marcos regulatorios, descubrimientos técnicos o experiencias de implementación.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

- Consulta con actores clave (organizaciones usuarias, expertos, comités éticos, etc.).

Este principio de evolución adaptativa garantiza que el sello mantenga su relevancia en contextos cambiantes, y actúe como herramienta viva de ética aplicada, no como una norma inmutable.

4.7 Consideraciones éticas y contextuales

El diseño de un sistema de certificación ética en inteligencia artificial aplicada a recursos humanos no puede limitarse a un marco técnico o normativo. Requiere, además, una comprensión profunda del contexto cultural, organizativo y social en el que se inserta, así como un compromiso real con los valores que pretende representar. Esta sección aborda tres riesgos críticos que deben gestionarse para preservar la legitimidad y eficacia del modelo propuesto.

4.7.1 Evitar la banalización de la ética: el riesgo del certification washing

Uno de los riesgos más relevantes en la implementación de un sello ético es su posible utilización como instrumento meramente simbólico o reputacional, sin un compromiso real por parte de la organización. Este fenómeno, conocido como certification washing, se refiere a la adopción superficial de sellos éticos como estrategia de imagen en lugar de como motor de transformación organizacional.

Organismos como el AI Now Institute han alertado sobre este tipo de prácticas, subrayando que el despliegue de una ética cosmética puede deslegitimar los esfuerzos reales y restar credibilidad a todo el ecosistema de certificación. Tal y como advierte Floridi en su análisis sobre los riesgos de una ética digital ineficaz, el simbolismo ético sin acciones reales puede convertirse en una forma de deslegitimación social (Floridi, L., 2019). La ética no debe prestarse al maquillaje institucional: tiene que operar como instrumento estructural de legitimidad y mejora organizativa.

Para mitigar este riesgo, el sistema de certificación debe:

- Establecer **requisitos proporcionados pero exigentes,** adecuados a cada nivel.
- Garantizar la **verificación independiente y trazable** de las evidencias.
- Incluir mecanismos de **revocación o suspensión** del sello ante incumplimientos.
- Fomentar la **publicación voluntaria de resultados,** como forma de transparencia responsable.

4.7.2 Evitar la sobrecarga normativa: equilibrio entre rigor y accesibilidad

Un segundo riesgo es la hiper regulación técnica, que puede hacer del modelo una herramienta excluyente, especialmente para pequeñas organizaciones, administraciones locales o entidades con menos recursos. Exigir un mismo nivel de cumplimiento a todas las organizaciones, sin contemplar su punto de partida ni su capacidad operativa, puede generar desigualdades, desincentivar la adopción o provocar efectos paralizantes.

Para asegurar el equilibrio entre ambición ética y viabilidad operativa, se recomienda:

- Aplicar una lógica **graduada y escalable** en los requisitos del sello.
- Mantener un enfoque **modular por procesos,** que permita certificar por etapas.

- Proporcionar herramientas de apoyo estandarizadas (guías, plantillas, ejemplos).
- Emplear un **lenguaje accesible**, comprensible más allá del entorno técnico o jurídico.

Estas medidas, en línea con los principios de gobernanza flexible del OECD AI Policy Observatory, favorecen la inclusión de todo tipo de organizaciones y la democratización del acceso a marcos de gobernanza algorítmica.

4.7.3 Contextualizar la ética: sensibilidad cultural, sectorial y organizativa

Por último, es fundamental evitar que la ética aplicada a la IA se convierta en una propuesta normativa universal sin sensibilidad al contexto. Las organizaciones operan en entornos culturales, normativos y de valores diversos, y los principios éticos deben poder adaptarse sin perder consistencia.

Estudios como el de Jobin, Ienca y Vayena han evidenciado que muchas guías éticas globales fracasan en su implementación precisamente por no considerar las dinámicas reales de las organizaciones (Jobin, 2019).

Para evitar este desajuste entre principios y práctica, el modelo de certificación debe:

- Permitir evaluaciones adaptadas al rol y al impacto del sistema en cada contexto.
- Incluir mecanismos de participación de los grupos afectados, desde empleados hasta sindicatos o comités sectoriales.
- Favorecer espacios de deliberación ética, más allá de la mera aplicación de normas.
- Promover la articulación de la ética como un proceso reflexivo y plural, en línea con el enfoque del European Group on Ethics in Science and New Technologies (EGE – Ethics of AI, 2021).

En definitiva, la certificación ética en IA aplicada a RR. HH. debe encontrar el equilibrio entre universalidad y contextualización, entre consistencia normativa y adaptación práctica, entre el deber ético y la viabilidad organizativa.

4.8 Rol de las administraciones públicas en la promoción de la certificación ética en IA

La consolidación de un ecosistema de inteligencia artificial ético y confiable en el ámbito de Recursos Humanos requiere no solo del compromiso voluntario de las organizaciones, sino también de un marco institucional proactivo que estimule, acompañe y legitime estos procesos. En este sentido, las administraciones públicas —a nivel local, autonómico, estatal y europeo— tienen un papel clave como facilitadoras y catalizadoras del modelo de certificación propuesto.

Funciones clave que podrían desempeñar:

- **Incorporar el sello como referencia en compras y licitaciones públicas**, especialmente cuando se contraten servicios de gestión de personal apoyados en IA, garantizando que los sistemas utilizados respeten criterios éticos verificables.

- **Promover políticas de incentivos** (económicos, fiscales, de visibilidad o acceso preferente a

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

convocatorias) para aquellas organizaciones que obtengan niveles avanzados de certificación, especialmente pymes.

- **Desarrollar programas de acompañamiento público,** como líneas de asesoramiento, mentorización y formación técnica para apoyar a las organizaciones en el proceso de implementación del sistema de certificación.

- **Fomentar la interoperabilidad normativa,** promoviendo la armonización del sello con otras normas ya existentes (calidad, protección de datos, compliance, igualdad, sostenibilidad…).

- **Impulsar alianzas público-privadas** para la creación de comunidades de práctica, espacios de experimentación (sandboxes), y observatorios que documenten el aprendizaje colectivo y los impactos del modelo.

- **Dar ejemplo como entidades empleadoras,** adoptando ellos mismos sistemas de certificación ética cuando implementen tecnologías de IA en la gestión de personas dentro de sus propias estructuras.

El compromiso activo de las administraciones no solo aportaría legitimidad y escala al modelo, sino que facilitaría su extensión homogénea y su impacto estructural, reforzando el valor estratégico de la ética en la transformación digital del empleo público y privado.

4.9 Conclusiones

La inteligencia artificial aplicada a la gestión de personas representa una de las fronteras más sensibles y estratégicas del desarrollo tecnológico contemporáneo. Su impacto directo sobre trayectorias profesionales, procesos de toma de decisiones y condiciones laborales exige mecanismos de gobernanza sólidos, verificables y adaptativos, capaces de garantizar que la tecnología sirva al bienestar y la equidad en el entorno de trabajo.

Este capítulo ha presentado un modelo de certificación ética en IA para RR. HH. basado en una lógica progresiva, estructurada y contextualizada, que permita a las organizaciones:

- Reconocer su nivel de madurez en materia de ética algorítmica.
- Identificar áreas de mejora y trazar itinerarios viables hacia la excelencia.
- Construir confianza con empleados, candidatos y grupos de interés.
- Contribuir a un ecosistema más justo, transparente y responsable.

1.Un sistema escalonado con base ética y técnica

La propuesta se articula a partir de:

- **Principios éticos adaptados al entorno laboral,** como la justicia, la explicabilidad, la no discriminación y la supervisión humana.
- Una **arquitectura por niveles (Bronce, Plata, Oro, Platino),** inspirada en modelos de madurez organizativa, que facilita una adopción gradual y medible.
- La definición de **elementos clave de evaluación** centrados en los datos, los modelos, la gobernanza, el impacto organizativo, la participación de terceros y la accesibilidad del sistema.

• Recomendaciones diferenciadas por tipo de organización (grandes empresas, pymes, sector público), junto con recursos de acompañamiento y mecanismos de verificación internos y externos.

2. Retos clave a considerar

En el diseño y despliegue de este sistema, se identifican tres retos que deben ser gestionados de forma crítica:

1. Evitar el certification washing: el sello no debe convertirse en un símbolo vacío, sino en una evidencia tangible de mejora continua, con revisión periódica y posibilidad de revocación.
2. Evitar la hiper regulación excluyente: el modelo debe mantener su exigencia sin convertirse en una barrera para organizaciones con menor capacidad técnica, facilitando su implementación mediante modularidad y apoyo.
3. Asegurar la contextualización ética: cada certificación debe adaptarse al contexto sectorial, cultural y organizativo, sin renunciar a su coherencia normativa.

3. Hacia una comunidad certificadora intersectorial

Más allá de su aplicación individual, el modelo de certificación propuesto aspira a consolidar una comunidad intersectorial de entidades comprometidas con la ética algorítmica. Esta comunidad podría:

• Compartir buenas prácticas, aprendizajes y experiencias de implementación.
• Impulsar el desarrollo de estándares normativos específicos para RR. HH.
• Alimentar un sistema de mejora colectiva y aprendizaje compartido.

Este enfoque colaborativo responde a la necesidad de democratizar la gobernanza de la IA en el trabajo, generando estándares que sean tanto exigentes como accesibles, tanto técnicamente sólidos como socialmente responsables.

4. La certificación como herramienta y no como fin

En definitiva, el sistema de certificación ética en IA para RR. HH. debe entenderse no como un punto de llegada, sino como una herramienta de cultura, transformación y rendición de cuentas. Su valor no radica solo en el cumplimiento de requisitos, sino en su capacidad para:

• Fomentar una cultura organizativa ética.
• Visibilizar avances concretos en el uso responsable de la tecnología.
• Promover un ecosistema laboral centrado en las personas, no en los algoritmos.

En este sentido, el sello ético propuesto puede convertirse en una palanca clave para la gobernanza inclusiva de la IA en el trabajo, guiando a las organizaciones hacia un futuro donde la inteligencia artificial no sea solo eficiente, sino también justa, confiable y humanamente significativa.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Apéndice del Capítulo 4

Hacia una certificación específica para la IA en RR. HH.

Las definiciones y estructuras aquí propuestas se inspiran en marcos internacionales (ISO, UNE, RIA), pero han sido adaptadas con lenguaje propio y enfoque sectorial para el contexto de RR. HH.

Hacia una certificación específica para IA en RR. HH. [11]: las definiciones y estructuras aquí propuestas se inspiran en marcos internacionales (ISO, UNE, RIA), pero han sido adaptadas con lenguaje propio y enfoque sectorial para el contexto de RR. HH.

Propuesta de norma técnica sectorial

Título propuesto: Norma para la gestión ética y certificable de sistemas de inteligencia artificial en procesos de Recursos Humanos (Posible denominación futura: UNE-ISO/IEC XXX- RR. HH.: Inteligencia Artificial Ética en la gestión de personas).

1. Objeto y campo de aplicación

La presente norma establece los requisitos, principios y estructuras necesarios para el diseño, implementación, evaluación y certificación ética de sistemas de inteligencia artificial aplicados a procesos de RR. HH. Esta propuesta se formula como base anticipatoria para una futura norma técnica sectorial, adaptada específicamente al uso de IA en contextos laborales.

La norma tiene por objeto:

- **Proporcionar una estructura certificable** que garantice el uso justo, trazable, explicable, accesible y no discriminatorio de sistemas de IA en el ámbito de la gestión de personas.
- **Prevenir impactos adversos** sobre los derechos fundamentales, la equidad de oportunidades, la autonomía profesional y la dignidad en el trabajo.
- **Facilitar la convergencia** con el Reglamento de Inteligencia Artificial de la Unión Europea y con normas internacionales en materia de gestión, auditoría, seguridad, ética digital y capital humano.

---

[11] **Nota**: Dado el vacío actual en las normas ISO y UNE específicas para la inteligencia artificial en el ámbito de los recursos humanos —a pesar de tratarse de un uso reconocido como de alto riesgo por el Reglamento Europeo de Inteligencia Artificial (RIA)—, este Libro Blanco incorpora, como ejercicio propositivo, una propuesta de norma técnica sectorial.

Este añadido no tiene carácter normativo oficial, pero se presenta como una base de trabajo anticipatoria, estructurada según la lógica de las normas internacionales, con el objetivo de:

- Servir de referente para futuras iniciativas de normalización (UNE, ISO, CEN, etc.).
- Orientar a organizaciones y entidades certificadoras en el desarrollo de sus propios esquemas éticos y de calidad.
- Fomentar una madurez compartida en el uso de IA en los procesos de gestión de personas.

La propuesta aquí recogida complementa el sistema de certificación progresiva descrito en el capítulo 5, alineándose con marcos existentes, pero también proponiendo criterios específicos, adaptados al impacto diferencial de la IA en el ámbito laboral.

• **Promover la madurez organizativa** en el uso de tecnologías algorítmicas, fomentando una cultura ética de gobernanza responsable de la IA en el trabajo.

Esta norma es aplicable a:

• Organizaciones públicas o privadas, de cualquier tamaño o sector, que desarrollen o utilicen sistemas de IA en procesos de toma de decisiones vinculados a la gestión de personas (por ejemplo, selección, promoción, formación, evaluación del desempeño, desvinculación o planificación de carrera).
• Terceros responsables del desarrollo, provisión, auditoría o certificación de dichos sistemas, incluyendo proveedores tecnológicos, consultoras, organismos certificadores, universidades y comités éticos sectoriales.

En particular, esta norma está orientada a ofrecer:

• **Un marco común de referencia** para evaluar la adecuación ética y técnica de sistemas de inteligencia artificial en RR. HH.
• **Una base operativa para esquemas de certificación escalables**, basados en niveles de madurez y enfoques modulares por proceso.
• **Una herramienta para entidades regulatorias,** que facilite la evaluación del cumplimiento de requisitos en el contexto del RIA, especialmente en relación con sistemas de IA clasificados como de alto riesgo.

Dado el carácter pionero de esta propuesta, su contenido deberá ser revisado, ampliado y contrastado progresivamente, a medida que se consoliden tanto la normativa europea como las prácticas sectoriales derivadas de su aplicación.

2. Términos y definiciones

Aunque hay un anexo específico con el glosario del Libro Blanco, nos parece pertinente, a efectos de esta norma, explicar los siguientes términos y definiciones utilizados en este. Todos ellos han sido seleccionados para garantizar la claridad conceptual, la coherencia normativa y la alineación con los marcos internacionales más relevantes en materia de inteligencia artificial y Recursos Humanos.

Inteligencia Artificial: Sistema basado en tecnologías computacionales diseñado para operar con grados variables de autonomía, que puede analizar datos, inferir patrones, tomar decisiones o recomendar acciones que afectan a entornos o personas (Adaptado de ISO/IEC 22989:2023).

Sistema de IA en RR. HH.: Conjunto de algoritmos, modelos, datos y procesos automatizados —como algoritmos, datos y procedimientos— que contribuyen en la toma de decisiones en áreas clave de recursos humanos, incluyendo selección, evaluación, promoción, formación, desarrollo y desvinculación de personas.

Gestión ética de la inteligencia artificial: Conjunto de principios, prácticas y controles que buscan asegurar que el uso de IA respete los derechos fundamentales, promueva la equidad, evite la discriminación, garantice la transparencia y permita el control humano significativo en los procesos decisionales.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Certificación algorítmica: Proceso estructurado que permite a una entidad acreditar, frente a terceros, que sus sistemas de IA cumplen con criterios éticos, técnicos organizativos y regulatorios.

Sesgo algorítmico: Resultado parcial, distorsionado o injustificado que se produce en un sistema de IA y que genera diferencias significativas o discriminatorias en el tratamiento de individuos o grupos, debido a errores en los datos, el diseño del modelo o su implementación.

Supervisión humana significativa: Nivel de control humano que asegura que las personas pueden entender, evaluar críticamente e intervenir en decisiones automatizadas, especialmente cuando tienen consecuencias relevantes en el entorno laboral.

Impacto ético-laboral: Efectos potenciales o reales que la aplicación de sistemas de IA puede generar sobre aspectos fundamentales como la autonomía individual, la justicia organizativa, la igualdad de oportunidades, el bienestar psicosocial o la dignidad profesional de las personas trabajadoras.

3. Estructura del sistema de gestión ética de IA en RR. HH.

Esta sección establece los elementos estructurales mínimos que una organización debe implementar para asegurar una gestión ética, trazable y certificable de sistemas de inteligencia artificial en el ámbito de los recursos humanos. La estructura se alinea con principios de mejora continua, gobernanza transversal, transparencia y rendición de cuentas, en coherencia con la ISO/IEC 42001, la ISO/IEC 27001 y otras normas de sistemas de gestión.

3.1 Política de IA ética en RR. HH.

Toda organización debería contar con una política institucional clara que exprese su compromiso con un uso ético y responsable de la inteligencia artificial en la gestión de personas. Esta política deberá:

- Incluir **principios éticos adoptados** (justicia, explicabilidad, no discriminación, trazabilidad, control humano significativo).
- Estar **aprobada por la alta dirección** y difundida a todos los niveles organizativos.
- Estar **sujeta a revisión periódica**, adaptándose a nuevos desarrollos normativos, tecnológicos o contextuales.

3.2 Gobierno del sistema de IA

Se recomienda que la organización defina una estructura de gobernanza transversal para la supervisión ética del sistema, que contemple:

- Asignación clara de **roles y responsabilidades** (perfiles técnicos, legales, éticos y de RR. HH.).
- Creación de **mecanismos internos de supervisión y revisión ética**, como por ejemplo comités interdisciplinares o responsables de cumplimiento.
- Enlace formal con **áreas de compliance, protección de datos, relaciones laborales y transformación digital.**

Esta estructura deberá estar documentada, con procedimientos de coordinación y toma de decisiones interdepartamentales.

3.3 Ciclo de vida del sistema algorítmico

Es esencial que la organización documente todo el recorrido del sistema de IA, desde su concepción hasta su retirada. Cada etapa —definición de objetivos, selección de proveedores, formación de usuarios, seguimiento, evaluación de impacto— debe quedar registrada y sujeta a revisión continua.

- **Fase de diseño:** definición del objetivo, evaluación de riesgos éticos y funcionales.
- **Fase de desarrollo o adquisición:** validación del proveedor, revisión de condiciones técnicas y jurídicas.
- **Fase de validación:** validación del sistema, prueba piloto con supervisión y documentación del proceso de validación y decisiones tomadas.
- **Fase de despliegue:** formación de usuarios, documentación accesible, pruebas de equidad.
- **Fase de uso:** seguimiento de resultados, auditorías internas, control de sesgos.
- **Fase de actualización o retirada:** justificación, trazabilidad, evaluación del impacto residual.

3.4 Formación y alfabetización algorítmica

La organización debe asegurar que todas las personas implicadas en el uso, interpretación o supervisión del sistema de IA cuenten con formación suficiente en:

- Principios de ética aplicada a la IA.
- Evaluación y mitigación de sesgos y riesgos.
- Derechos digitales y laborales vinculados a decisiones automatizadas con supervisión humana.
- Uso seguro y responsable del sistema.

Esta formación deberá adaptarse al rol (por ejemplo, técnico, directivo, evaluador, operador) y actualizarse periódicamente.

3.5 Participación de las personas afectadas

El sistema debe prever mecanismos claros para garantizar la información, participación y derecho de revisión por parte de las personas trabajadoras cuyos datos o trayectorias puedan verse afectadas por la inteligencia artificial:

- Comunicación previa sobre el uso de sistemas de IA en procesos decisionales relevantes.
- Acceso a explicaciones comprensibles sobre las decisiones.
- Posibilidad de ejercer el derecho a revisión, corrección o apelación.
- Participación en procesos de evaluación del sistema cuando el impacto sea significativo.

Este principio responde tanto al enfoque de supervisión humana significativa como a la necesidad de legitimidad organizativa y diálogo social.

4. Requisitos técnicos y éticos del sistema de IA en RR. HH.

Este apartado establece los requisitos que deben cumplir los sistemas de inteligencia artificial utilizados en el ámbito de los recursos humanos, con el fin de garantizar su alineación con principios éticos,

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

técnicos y legales. Estos requisitos son aplicables tanto a sistemas desarrollados internamente como a soluciones adquiridas externamente, y deben ser verificables en el marco del proceso de certificación.

4.1 Legalidad y licitud del tratamiento de datos

El sistema de IA debe operar sobre datos cuyo tratamiento cumpla con la legislación vigente en materia de protección de datos personales, privacidad y seguridad de la información.

- Los datos tienen que ser **obtenidos legalmente**, pertinentes y adecuados para la finalidad prevista.
- Es necesario aplicar **medidas de minimización, anonimización y conservación limitada**, según lo previsto en el RGPD y estándares como ISO/IEC 27701:2019.
- Hay que garantizar la **actualización y calidad de los datos** para evitar sesgos derivados de errores u obsolescencia.

4.2 No discriminación y equidad algorítmica

La organización debe demostrar que ha aplicado mecanismos adecuados para prevenir, detectar y corregir cualquier forma de sesgo injustificado.

- Se prohíbe la discriminación directa o indirecta por razón de sexo, edad, etnia, discapacidad, orientación sexual u otras condiciones personales o sociales.
- El sistema debe ser sometido a evaluaciones de sesgo algorítmico durante su entrenamiento, despliegue y uso, siguiendo principios de la ISO/IEC TR 24027:2023.
- Las medidas de corrección deben estar documentadas y aplicarse de forma proactiva, sin esperar a la identificación de un perjuicio.

4.3 Transparencia y explicabilidad

Los sistemas de IA deben garantizar niveles suficientes de transparencia y explicabilidad, ajustados al perfil de cada parte implicada.

- Se debe proporcionar **información clara, accesible y verificable** sobre el funcionamiento del sistema, incluyendo variables clave, criterios de decisión y fuentes de datos.
- Deben implementarse **mecanismos de explicabilidad adaptada** a diferentes tipos de público: técnicos, auditores, usuarios operativos y personas afectadas.
- Todas las decisiones o recomendaciones automatizadas con supervisión humana deben estar **documentadas**, incluyendo su trazabilidad lógica y procedimental.

4.4 Supervisión humana significativa

Las decisiones automatizadas deben estar sujetas a control humano efectivo y verificable, garantizando la posibilidad de intervención en todas las fases.

- Las personas designadas deben estar **formadas, autorizadas y capacitadas** para interpretar y corregir los resultados del sistema.
- En decisiones de alto impacto (ej. contratación, promoción, despido), debe existir siempre la **posibilidad de revisión o anulación humana.**

• No se permitirá que el sistema sustituya funciones críticas del juicio humano sin justificación clara y documentada.

4.5 Evaluación de impacto ético-laboral

Antes del despliegue o actualización sustancial del sistema, la organización debe realizar una **evaluación de impacto ético-laboral**, que incluya:

• Identificación de riesgos potenciales sobre la equidad, la autonomía y la dignidad profesional.
• Medidas de mitigación y planes de contingencia ante escenarios de riesgo alto.
• Consulta y participación de los grupos potencialmente afectados, como pueden ser los empleados, representantes sindicales o comités sectoriales.

Esta evaluación deberá quedar documentada y ser accesible para revisión ética o auditoría externa.

5. Mecanismos de certificación y verificación externa

Este apartado establece los procedimientos que permiten verificar, de forma objetiva e independiente, el cumplimiento de los requisitos técnicos, éticos y organizativos establecidos en esta norma. La certificación constituye una verificación formal, escalonada por niveles, orientada a fomentar la mejora continua y la confianza social en el uso responsable de sistemas de IA aplicados a Recursos Humanos.

5.1 Certificación por niveles

La certificación se articula en cuatro niveles progresivos, basados en un modelo de madurez ética y operativa. Cada nivel supone un umbral creciente de exigencia y compromiso organizativo.

• **Nivel Bronce – Compromiso inicial:** adopción formal de principios éticos, realización de una autoevaluación estructurada y diseño de un plan de mejora inicial.
• **Nivel Plata – Implementación activa:** evidencias de acciones correctoras, formación del personal implicado, existencia de protocolos éticos internos y auditoría externa satisfactoria.
• **Nivel Oro – Excelencia ética:** trazabilidad completa del sistema, participación activa de terceros, gobernanza sólida, explicabilidad validada y cumplimiento de todas las dimensiones evaluables.
• **Nivel Platino – Liderazgo sectorial:** contribución a estándares sectoriales, transparencia pública de procesos y resultados, innovación ética, publicación de evaluaciones de impacto y colaboración con otras entidades.

Cada nivel podrá ir acompañado de una declaración de conformidad formal, firmada por la entidad certificadora y sujeta a revisión periódica.

5.2 Auditorías externas independientes

Las organizaciones que aspiren a la certificación deberán someterse a una auditoría ética y técnica realizada por un tercero acreditado. Esta auditoría debe contemplar:

• Revisión completa de la documentación técnica, organizativa y formativa del sistema.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

• Entrevistas con los distintos perfiles implicados: responsables técnicos, usuarios operativos, empleados afectados, etc.
• Verificación de la trazabilidad algorítmica, decisiones automatizadas y medidas de mitigación aplicadas.
• Validación del cumplimiento de los requisitos establecidos en cada dimensión (datos, modelo, impacto, participación, accesibilidad…).

Las auditorías podrán utilizar marcos como ISO 19011:2018 (auditorías de sistemas de gestión) o ISO/IEC 27007:2020 (auditoría de seguridad de la información), y se adaptarán al nivel de certificación solicitado.

5.3 Declaración de conformidad y vigencia del sello

Superada la auditoría, la entidad podrá obtener una declaración de conformidad acompañada del sello correspondiente al nivel alcanzado. La vigencia del sello será limitada y se regirá por los siguientes plazos:

• **2 años** para los niveles Bronce y Plata.
• **3 años** para los niveles Oro y Platino.

Durante ese período, se exigirá una revisión intermedia simplificada (tipo seguimiento), que permita verificar la continuidad del compromiso. En caso de incumplimientos graves, la certificación podrá ser suspendida o revocada, con obligación de retirada del sello.

5.4 Mecanismos de revisión, suspensión o revocación

El sistema de certificación incluirá medidas formales para preservar su legitimidad y credibilidad:

• **Revisión anticipada** en caso de incidentes éticos, cambios sustanciales en el sistema o denuncias fundamentadas.
• **Suspensión temporal** cuando se identifiquen desviaciones corregibles.
• **Revocación definitiva** ante incumplimientos reiterados o falta de cooperación con el proceso de verificación.
• Derecho de **apelación** por parte de la organización auditada ante una comisión técnica independiente.

Este enfoque garantiza que el sello no funcione como una certificación estática, sino como un instrumento dinámico de responsabilidad ética y trazabilidad continua.

6. Correspondencias normativas y marco de integración

Este apartado presenta un mapa de normas internacionales, guías técnicas y marcos regulatorios relevantes que se vinculan directamente con la presente propuesta de norma sectorial.

Su finalidad es facilitar la compatibilidad técnica, jurídica y organizativa del sistema de certificación, así como maximizar su aplicabilidad en auditorías, procesos de cumplimiento y esquemas de gobernanza de la IA.

6.1 Normas ISO/UNE de aplicación directa o complementaria

Tabla 25: *Ámbito de aplicación de las normas ISO/UNE*

Norma	Ámbito	Relación con esta norma
ISO/IEC 42001:2023	Gestión de sistemas de IA	Marco general de gestión aplicable. Esta norma sectorial actúa como especialización.
ISO/IEC 23894:2023	Gestión de riesgos en IA	Apoya los requisitos de evaluación de riesgos y medidas de mitigación.
ISO/IEC TR 24027:2023	Evaluación de sesgos algorítmicos	Base para pruebas de equidad y auditoría técnica del modelo.
ISO/IEC TR 24368:2022	Consideraciones éticas y sociales en IA	Referente transversal para el diseño ético y la supervisión humana.
ISO/IEC 22989:2023	Terminología de IA	Proporciona coherencia conceptual con el vocabulario normativo internacional.
ISO/IEC 27001:2022	Seguridad de la información	Garantiza la confidencialidad, integridad y disponibilidad de los datos tratados.
ISO/IEC 27701:2019	ISO/IEC 27701:2019	Extensión de ISO 27001 para proteger información personal en entornos algorítmicos, especialmente relevante en contextos con datos sensibles.
ISO 30414:2018	Información sobre capital humano	Complementa la trazabilidad ética con indicadores de RR.HH..
ISO 19011:2018	Directrices para auditorías de sistemas de gestión	Referencia para la estructura de auditoría y revisión interna del sistema.
ISO/IEC 27007:2020	Auditoría de sistemas de seguridad de la información	Útil para evaluar trazabilidad, confidencialidad y protección de datos.
ISO 37301:2021	Sistemas de gestión del compliance	Refuerza los aspectos organizativos y normativos del sistema propuesto.

Fuente: elaboración propia.

6.2 Relación con el Reglamento europeo de Inteligencia Artificial

La presente norma se alinea con el RIA, particularmente en lo relativo a los sistemas de IA de alto riesgo, como es el caso de los sistemas aplicados a RR. HH. En particular:

- Reconoce el carácter de **alto impacto y sensibilidad** de estos sistemas sobre los derechos de las personas.
- Integra principios clave del Reglamento europeo de IA, como **la supervisión humana significativa, la trazabilidad, la documentación técnica adecuada y la transparencia del funcionamiento algorítmico.**
- Ofrece una **base organizativa y técnica para demostrar el cumplimiento proactivo** ante autoridades competentes, tanto en auditorías internas como en procesos de conformidad externa.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

6.3 Finalidad de este apéndice

Este apartado no pretende ser exhaustivo, sino operativo y orientador. Se reconoce que la norma sectorial aquí propuesta deberá evolucionar en los próximos años para:

- Integrar nuevas referencias normativas a medida que se publiquen normas ISO específicas sobre IA en sectores clave (salud, educación, administración pública, etc.).
- Alinearse con futuros desarrollos del RIA y con prácticas emergentes en auditoría algorítmica.
- Incorporar lecciones aprendidas durante la aplicación práctica del sistema en distintos tipos de organizaciones.

5. RETOS FUTUROS Y CONCLUSIONES

Este capítulo recoge las conclusiones del Libro Blanco sobre la aplicación de la inteligencia artificial en el ámbito de Recursos Humanos y los retos futuros que ya están en marcha. Se basa en un análisis empírico construido a partir del estudio de casos reales proporcionados por organizaciones que han colaborado de forma voluntaria y desinteresada. La riqueza del trabajo reside en la diversidad de estos casos y en la implicación activa de los equipos participantes.

Las conclusiones aquí presentadas han sido estructuradas por aspectos temáticos, alineadas con los objetivos del proyecto, para ofrecer una lectura accesible a perfiles diversos: desde responsables políticos y directivos, hasta expertos técnicos, consultores o miembros del público general interesado en los retos de la IA en el entorno laboral. Además, se incluye una tabla resumen final con los principales hallazgos, recomendaciones y su vinculación con los tres ejes estratégicos del proyecto.

5.1 El nuevo mercado laboral

El mercado laboral contemporáneo se encuentra en transformación acelerada por el avance de la inteligencia artificial, que redefine no sólo los procesos de selección y evaluación, sino también las competencias requeridas, la trazabilidad de los perfiles y la estructura misma del empleo.

La aplicación de IA en este contexto ha puesto de manifiesto la existencia de sesgos estructurales en los algoritmos utilizados, que afectan directamente la empleabilidad de colectivos vulnerables. A menudo, estos sistemas están diseñados sin tomar en cuenta la complejidad del entorno humano-laboral, lo que puede generar efectos excluyentes y reforzar dinámicas preexistentes de desigualdad.

Asimismo, se han detectado riesgos emergentes como la fragmentación del trabajo, la hiper segmentación de perfiles o la automatización de tareas sin análisis previo del impacto social y organizativo. Frente a ello, se vuelve urgente promover una visión ética del desarrollo tecnológico, que priorice la justicia social, la equidad en el acceso al empleo y la protección de los derechos de los trabajadores.

Además, el uso de IA ha puesto en evidencia una falta de preparación generalizada por parte de los empleadores para entender los riesgos y oportunidades que estas herramientas implican. Muchos adoptan soluciones sin evaluar su impacto, sin contar con criterios claros para seleccionar proveedores o sin garantizar el derecho a la explicabilidad de las decisiones tomadas a partir de resultados automatizados.

En este nuevo mercado laboral, es imprescindible:

- Entender que **no toda innovación tecnológica es neutral ni beneficiosa por defecto.**
- Asegurar que la IA se despliega con **perspectiva inclusiva, legal y estratégica.**
- Fortalecer la **capacidad crítica de las organizaciones** en torno al uso de datos y algoritmos.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

5.2 Escalabilidad de las soluciones planteadas

Uno de los aprendizajes fundamentales del proyecto ha sido la necesidad de diseñar soluciones proporcionales y adaptables, que permitan a diferentes tipos de organizaciones —desde grandes corporaciones hasta pymes o entidades del tercer sector— abordar la aplicación de IA con garantías, sin quedar bloqueadas por la complejidad técnica o el miedo al incumplimiento normativo.

La propuesta de una certificación ética escalable e incremental se sitúa en el centro de esta visión. Esta herramienta permite modular el nivel de exigencia en función del grado de madurez digital, cultural y organizativa de cada entidad, estableciendo un recorrido claro y progresivo hacia una IA más ética, explicable y socialmente responsable.

El enfoque de escalabilidad responde a tres objetivos clave:

- **Reducir las barreras de entrada:** evitando que únicamente las grandes empresas puedan implementar IA de forma segura y regulada.
- **Fomentar la mejora continua:** ofreciendo un marco que permita avanzar desde niveles básicos de cumplimiento hasta niveles avanzados de transparencia y gobernanza.
- **Construir confianza:** tanto a nivel interno (equipos, responsables) como externo (usuarios, personas candidatas, ciudadanía en general).

Esta visión implica no solo definir estándares, sino también generar capacidades institucionales y formativas que conviertan la ética algorítmica en una práctica cotidiana, comprensible y aplicable. Para ello, resulta clave desplegar una infraestructura pública de apoyo que incluya marcos orientativos, servicios de asesoramiento, bancos de buenas prácticas, plantillas de evaluación de impacto y recursos accesibles que impulsen la equidad tecnológica sin sobrecargar a las organizaciones más pequeñas.

5.3 Retos legales futuros en un contexto de IA avanzada

A medida que la inteligencia artificial evoluciona hacia formas más complejas y autónomas, los desafíos legales se intensifican. Las normativas actuales —aunque han supuesto un gran avance— aún no cubren con precisión las nuevas configuraciones tecnológicas que se consolidan en los entornos laborales. Entre estos desafíos emergentes destacan cinco áreas prioritarias:

1. Regulación de modelos auto aprendientes

Los algoritmos que modifican su comportamiento en tiempo real (modelos autoajustables o auto aprendientes) plantean un gran dilema jurídico: ¿cómo asegurar que un modelo que evoluciona de forma continua mantiene el cumplimiento normativo inicial? Esto requiere marcos de validación continuada, auditorías vivas y mecanismos legales que reconozcan la naturaleza dinámica de estas tecnologías.

2. Estandarización de auditorías algorítmicas

Hoy en día no existe una metodología homogénea para evaluar algoritmos en el ámbito de los RR. HH. Se hace urgente establecer criterios técnicos, éticos y jurídicos estandarizados, que permitan evaluar comparativamente distintos sistemas, certificar su uso y generar confianza en su despliegue.

3. Gobernanza del dato en entornos federados

El machine learning federado y otros sistemas distribuidos de procesamiento de datos abren grandes oportunidades de protección de la privacidad, pero también hacen más compleja la atribución de responsabilidades y la garantía de los derechos digitales. Será necesario desarrollar nuevos enfoques de gobernanza que reconozcan estas arquitecturas descentralizadas.

4. Responsabilidad en IA generativa aplicada a RR. HH.

La expansión de herramientas generativas (que redactan informes, diseñan entrevistas, elaboran perfiles de talento, etc.) introduce una nueva dimensión en la toma de decisiones tomadas a partir de resultados automatizados. Será esencial definir quién responde por la información generada, cómo se valida y en qué condiciones puede formar parte de un proceso formal de selección o evaluación.

5. Transparencia avanzada y explicabilidad en algoritmos de segunda generación

Ya no solo hablamos de IA que toma decisiones directamente sobre personas, sino de meta-algoritmos: IA que supervisa, pondera o corrige los resultados de otras IA. Estos escenarios exigen nuevos niveles de explicabilidad y trazabilidad, así como marcos legales específicos para el control de la delegación entre sistemas automatizados.

Además, comienzan a consolidarse marcos normativos que refuerzan los derechos de información de los representantes de los trabajadores sobre el uso de sistemas automatizados en la gestión laboral. El artículo 64.4.d) del Estatuto de los Trabajadores, reforzado por la Disposición Adicional 23.ª del Real Decreto-ley 9/2021 (Ley Rider), obliga a las empresas a informar sobre el funcionamiento de sistemas algorítmicos que afecten a la organización del trabajo, la asignación de tareas o la evaluación del desempeño. Esta línea de desarrollo apunta a un modelo donde la transparencia algorítmica será exigible no sólo hacia las personas trabajadoras, sino también hacia sus representantes legales, estableciendo un nuevo estándar de gobernanza digital laboral.

5.4 Conclusiones finales

Tras el análisis del nuevo escenario laboral, los requisitos para una escalabilidad viable de soluciones y los retos legales emergentes, se consolidan a continuación las principales conclusiones del proyecto. Estas se presentan organizadas por ejes temáticos, en línea con los objetivos estratégicos del Libro Blanco, y ofrecen una visión transversal que articula las evidencias empíricas con recomendaciones concretas para empresas, administraciones y agentes sociales.

Su propósito es servir como base para la acción y como punto de partida para futuras políticas, estándares técnicos y líneas de investigación en torno a una IA laboral más ética, inclusiva y eficaz.

1. Sesgos en el diseño y uso de algoritmos en función de los ejes de actuación

Eje 5: Inserción sociolaboral

Uno de los hallazgos más preocupantes del análisis es la presencia de sesgos no detectados en los sistemas analizados empíricamente. Muchas organizaciones, con la intención de evitar discriminaciones, eliminan variables como género o discapacidad. Sin embargo, esta práctica no

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030 Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

garantiza la neutralidad, sino que puede ocultar sesgos implícitos que se mantienen a través de otras variables correlacionadas.

Además, ninguna de las empresas evaluadas realizó un análisis técnico riguroso para identificar sesgos potenciales. Se evidencia una falta de comprensión sobre cómo los sesgos pueden surgir de forma estructural y cómo deberían abordarse de forma proactiva, más allá de la simple omisión de variables sensibles.

El impacto de estos sesgos es especialmente grave en colectivos vulnerables, cuya invisibilización puede comprometer directamente sus oportunidades de empleabilidad.

2. Cumplimiento normativo y Gobernanza del dato

Eje 3: Transformación digital

Los análisis empíricos revelan un cumplimiento muy deficiente de la normativa vigente. Ninguna de las organizaciones realizó una evaluación de impacto previa, a pesar de tratarse de algoritmos de alto riesgo en el marco de la regulación europea (RIA). Además, se han identificado usos no autorizados de algoritmos más allá de su finalidad original, lo que constituye una infracción grave del principio de "limitación de la finalidad".

Este vacío normativo se agrava por una falta de gobernanza interna del dato: las empresas no disponen de estructuras ni procedimientos claros para gestionar el ciclo de vida de los datos ni para supervisar las implicaciones legales de sus sistemas automatizados.

3. Formación y cultura organizativa

Eje 3: Transformación digital

Existe un fuerte desequilibrio entre la formación técnica (uso de herramientas como ChatGPT, Copilot, etc.) y la formación ética, crítica o estratégica. La mayoría de los empleados y directivos desconocen los riesgos reales asociados al uso de IA y carecen de recursos para evaluar adecuadamente su implementación.

Además, se observa una externalización excesiva del desarrollo de algoritmos, sin supervisión suficiente por parte de los responsables de RR. HH. Esto contribuye a una cultura de dependencia tecnológica sin control, donde las decisiones quedan en manos de los técnicos sin una visión de conjunto.

4. Inclusión social y colectivos invisibilizados

Ejes 1 y 5: Promoción de derechos e inclusión + Inserción sociolaboral

La eliminación de variables sensibles ha llevado a la invisibilización de colectivos vulnerables, como personas con discapacidad, migrantes o víctimas de violencia de género. Al no registrar datos sobre condiciones relevantes, se impide realizar análisis que identifiquen desigualdades estructurales.

Se proponen alternativas como el uso de datos sintéticos para representar estos perfiles de forma

segura. También destacan enfoques innovadores como describir "capacidades funcionales" en lugar de "discapacidades", lo que abre nuevas posibilidades para la inclusión desde el diseño algorítmico. Estas estrategias contribuyen directamente a la mejora de la empleabilidad, autonomía y protección de los derechos de estos colectivos.

5. RR. HH.: debilidades en la compra y uso de IA

Eje 5: Inserción sociolaboral

Desde la perspectiva de Recursos Humanos, se ha constatado una falta generalizada de preparación para evaluar, seleccionar e implementar sistemas de inteligencia artificial. No se realizan análisis de impacto, no se consideran alternativas menos invasivas, y no se exige a los proveedores el cumplimiento de criterios de explicabilidad o proporcionalidad.

El uso de modelos excesivamente complejos, sin justificación técnica clara, añade opacidad al proceso. En muchos casos, soluciones más simples y explicables habrían bastado. Esta tendencia refleja un desajuste entre el entusiasmo por la IA y la comprensión real de sus implicaciones prácticas.

6. Certificación ética e implementación progresiva

Ejes 3 y 5: Transformación digital + Inserción sociolaboral

Una conclusión clave es la necesidad de establecer mecanismos de certificación ética adaptados al grado de madurez de cada organización. Se propone un modelo escalable e incremental, que empiece con requisitos básicos como la existencia de formación específica, comités éticos internos y procesos mínimos de documentación.

Esta certificación debería fomentar una cultura ética desde la base, ofreciendo seguridad jurídica y metodológica a las organizaciones, y facilitando su avance hacia modelos de IA más seguros, explicables y socialmente responsables.

7. Reflexiones finales: desconocimiento y riesgo

Ejes 1, 3 y 5: Derechos, transformación e inserción

Muchas organizaciones aplican IA sin saber realmente cómo funciona ni qué consecuencias puede tener. Se utilizan modelos complejos innecesarios, difíciles de explicar, cuando en muchos casos bastarían métodos más simples y comprensibles.

Esta incertidumbre genera dos tipos de reacción: paralización (frenar la innovación por miedo) o inconsciencia (aplicar inteligencia artificial sin evaluación). Ambas respuestas son inadecuadas y comprometen la equidad, la legalidad y los derechos fundamentales.

En el ámbito de la innovación social, esto representa un triple reto:

- Cualificar a los profesionales para el uso ético de IA
- Reconocer a los colectivos vulnerables en los datos
- Impedir que la tecnología agrave la desigualdad o perjudique los ODS

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Tabla 26: *Resumen de los principales hallazgos del el trabajo de investigación empírica*

Eje temático	Problema observado	Conclusión principal	Recomendación	Ejes Proyecto
Sesgos	Eliminación de variables sensibles sin análisis de fondo	Persisten sesgos implícitos	Realizar auditorías de sesgos con enfoque estructural	Eje 5
Normativa y gobernanza	No se hacen evaluaciones de impacto ni se controla el uso de algoritmos	Incumplimiento sistemático del RIA y falta de gobernanza	Aplicar evaluaciones de impacto y establecer control interno de algoritmos	Eje 3
Formación	Formación técnica sin perspectiva crítica	Ausencia de cultura ética y estratégica	Integrar pensamiento crítico y formación ética en todos los niveles	Eje 3
Inclusión	Invisibilización de colectivos vulnerables	Se impide detectar desigualdades reales	Usar datos sintéticos y enfoques centrados en capacidades	Ejes 1 y 5
Recursos Humanos	Desconocimiento en la compra y aplicación de IA	RR. HH. no sabe evaluar ni supervisar algoritmos	Generar guías internas y formación para responsables de RR. HH.	Eje 5
Certificación	No hay criterios unificados ni escalabilidad	Falta de confianza y miedo generalizado	Promover certificación escalable desde la sensibilización hasta la auditoría	Ejes 3 y 5
Reflexiones finales	Desconocimiento del funcionamiento y riesgos de la IA	Uso inadecuado de la IA | Inmovilismo por miedo	Fomentar cultura de innovación responsable y evaluación multidisciplinar	Ejes 1, 3 y 5

Fuente: elaboración propia.

Estas conclusiones resumen los hallazgos clave del proyecto y constituyen una base para el diseño de políticas, formación, regulación y buenas prácticas que permitan una implantación de la IA en Recursos Humanos más justa, transparente y eficaz.

ANEXOS

Financiado por la Unión Europea
NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid
Dirección General de Evaluación, Calidad e Innovación
CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Introducción

Como se ha explicado en la Introducción, al detallar la metodología de elaboración de este Libro Blanco, ha sido decisiva la colaboración de numerosos profesionales que han aportado un enfoque multidisciplinar. Este Anexo de la Red Colaborativa pone el foco en algunos temas que han ido surgiendo en diferentes capítulos acerca de la implantación de la IA en Recursos Humanos: desde la cultura y la Gobernanza del dato, y la formación de los profesionales, hasta la normativa vigente y los principios éticos aplicables. Los autores son:

Consejo Asesor de IA+Igual

- José Domínguez: *Protección de datos e IA*
- Olga Herrero: *Reskilling y upskilling de los profesionales de RR. HH. para implantar con éxito la IA*
- Javier Mocoso: *Aspectos éticos del uso de la IA en RR. HH.*
- Ana Valera: *Cultura del dato en RR. HH.: de los indicadores de desempeño a los modelos de IA*

Embajadores

- Carlos Cabañas: *Cómo afecta la IA a los procesos del área de talento*
- Estela Martín: *Normativa laboral española específica en torno a la IA*

Colaboradores

- Joaquín García Onrubia: *La importancia de la Gobernanza del dato en la empresa*
- Cristina Martín: *Día a día de la IA en la gestión de personas*

A través del cruce entre el análisis técnico, jurídico y ético, se advierte sobre los riesgos de aplicar tecnologías sin control humano y sin un marco normativo y ético robusto. La necesidad de interdisciplinariedad, formación continua y transparencia algorítmica se presentan como fundamentos para la construcción de una IA que respete la dignidad humana y promueva entornos laborales justos e inclusivos.

Protección de datos e IA

José Domínguez, abogado y socio de Martínez-Echevarría Abogados.
Miembro del Consejo Asesor del Proyecto IA+Igual.

La relación entre inteligencia artificial y la protección de datos personales representa uno de los mayores desafíos para el derecho contemporáneo. Este fenómeno se manifiesta en sectores tan diversos como el Financiero, Sanitario, Educativo y el de Seguridad Pública, entre otros. En cada uno de estos sectores, las implicaciones del uso de algoritmos y modelos de inteligencia artificial que procesan datos personales requieren un escrutinio normativo riguroso y una implementación técnica responsable.

Por ejemplo, en el ámbito de la justicia predictiva, los sistemas de inteligencia artificial pueden influir en la concesión de libertades provisionales, imponiendo sesgos derivados de datos históricos (antecedentes), muchas veces afectados por desigualdades estructurales. Este tipo de decisiones debe ser auditado, documentado y sometido a una evaluación ética y jurídica continua.

1. Concepto y alcance de la IA

En los últimos años, la inteligencia artificial ha dejado de ser un concepto de ciencia ficción para convertirse en una realidad presente en nuestra vida y en nuestro día a día. Desde los asistentes virtuales que nos ayudan a organizar el día, hasta los sistemas que recomiendan qué serie ver o incluso los algoritmos que detectan enfermedades, la inteligencia artificial está revolucionando la forma en que vivimos, trabajamos y nos relacionamos. Pero ¿qué es exactamente la inteligencia artificial y cuál es su verdadero alcance?

¿Qué es la inteligencia artificial?

La inteligencia artificial es una rama de la informática que busca desarrollar sistemas capaces de realizar tareas que, hasta hace poco, solo podían hacer los seres humanos. Estas tareas incluyen el reconocimiento de voz, la comprensión del lenguaje, el aprendizaje, la toma de decisiones o la resolución de problemas.

En otras palabras, una IA es un conjunto de algoritmos diseñados para imitar, en mayor o menor medida, el comportamiento inteligente humano. No se trata solo de seguir instrucciones, como lo hace un programa tradicional, sino de aprender a partir de los datos, adaptarse y mejorar con el tiempo.

Tipos de inteligencia artificial

La IA puede clasificarse de diversas maneras, pero una de las más comunes distingue entre:

- **Débil** o **estrecha**: se enfoca en tareas específicas, como traducir idiomas, jugar al ajedrez o recomendar productos. Es la forma más común de IA y la que usamos a diario.

- **Fuerte** o **general**: busca crear máquinas con capacidades cognitivas similares a las del ser humano, capaces de razonar, entender el contexto y actuar con sentido común. Este tipo de IA todavía está en fase de investigación.

- **Superinteligencia artificial**: una forma hipotética de inteligencia artificial que superaría la inteligencia humana en todos los aspectos.

Aunque es un tema fascinante, sigue siendo objeto de debate filosófico y científico.

Aplicaciones actuales de la IA

El alcance de la inteligencia artificial es vasto y se extiende a múltiples campos:

• **Salud**: diagnóstico médico asistido por IA, análisis de imágenes, desarrollo de medicamentos, etc.

• **Educación**: plataformas personalizadas de aprendizaje, tutores virtuales, análisis de desempeño estudiantil…

• **Industria y Transporte:** robots en fábricas, vehículos autónomos y mantenimiento predictivo, entre otros muchos.

• **Entretenimiento y redes sociales**: distintas recomendaciones de contenido, filtros de imágenes, asistentes de escritura, etc.

• **Finanzas**: detección de fraudes, análisis de riesgo crediticio, inversiones automatizadas…

Desafíos y consideraciones éticas

A pesar de sus avances, la IA también plantea desafíos importantes. La privacidad de los datos, el sesgo en los algoritmos, la automatización del trabajo y la toma de decisiones automatizadas son cuestiones éticas que requieren atención.

Además, se hace cada vez más necesario establecer marcos legales y normativos que garanticen un desarrollo responsable de estas tecnologías, equilibrando innovación con derechos fundamentales.

2. IA y tratamiento de datos personales

En plena era digital, la inteligencia artificial se ha convertido en una herramienta poderosa que revoluciona industrias enteras. Desde asistentes virtuales hasta sistemas de salud predictivos, la IA necesita un ingrediente esencial para funcionar: datos. Y, entre esos datos, muchos son personales. Esto nos plantea una pregunta clave: ¿cómo se manejan nuestros datos personales en un mundo gobernado por algoritmos?

¿Qué relación hay entre IA y datos personales?

La inteligencia artificial se basa en el análisis de grandes volúmenes de información para "aprender" y tomar decisiones. Cuantos más datos tenga, mejor puede funcionar. Muchos de esos datos provienen de personas: nombres, ubicaciones, preferencias, historiales de navegación, información médica, entre otros.

Por ejemplo, un sistema de recomendación de películas aprende de tus gustos anteriores. Una aplicación de salud, de tus síntomas y hábitos. Sin embargo, este uso intensivo de datos también puede comprometer tu privacidad si no se gestiona adecuadamente.

¿Qué se considera un dato personal?

Según leyes como el Reglamento General de Protección de Datos (RGPD) en Europa o diversas leyes latinoamericanas de protección de datos, un dato personal es cualquier información que permita identificar directa o indirectamente a una persona. Esto incluye:

• Nombre y apellido
• Número de identificación
• Dirección IP
• Datos de localización
• Historial médico
• Fotos, audios, vídeos

Cuando una IA procesa estos datos, debe hacerlo bajo ciertas reglas éticas y legales.

Riesgos del uso de IA sin protección de datos

Aunque la IA puede generar beneficios enormes, también conlleva riesgos si no se regula bien su relación con los datos personales:

• **Pérdida de privacidad:** si no hay control sobre

cómo se usan los datos, terceros podrían acceder a información sensible sin consentimiento.

• **Discriminación algorítmica**: algunos sistemas de IA reproducen o incluso amplifican sesgos existentes si los datos de entrenamiento están mal seleccionados.

• **Toma de decisiones opaca**: muchas veces no se entiende cómo una IA llegó a una conclusión, lo que dificulta reclamar si se comete un error (por ejemplo, en un rechazo de crédito).

¿Cómo proteger los datos personales frente a la IA?

La buena noticia es que existen marcos legales y principios éticos para proteger tus datos:

1. Consentimiento informado: debes saber y aceptar que tus datos serán usados por sistemas de IA.

2. Finalidad específica: la IA debe usar tus datos solo para el propósito que se te informó.

3. Minimización de datos: únicamente deben recopilarse los datos necesarios.

4. Transparencia: se tienen que explicar las distintas decisiones automatizadas que te afecten significativamente.

5. Derecho al olvido y acceso: puedes pedir que se eliminen tus datos o saber qué información tienen sobre ti.

¿Hacia una IA más ética?

El desarrollo de una IA ética y responsable es uno de los grandes desafíos del siglo XXI. Gobiernos, empresas y organizaciones están trabajando en guías y normativas para garantizar que el uso de IA no atente contra los derechos fundamentales.

La clave está en el equilibrio: aprovechar el potencial de la IA sin sacrificar la privacidad y la dignidad humana. Porque si bien los algoritmos pueden ser inteligentes, el verdadero progreso radica en usarlos con humanidad.

3. Aplicación del RGPD a la IA

La evolución acelerada de la inteligencia artificial plantea nuevos retos en materia de privacidad y protección de datos. En Europa, el RGPD, en vigor desde mayo de 2018, se erige como la norma clave para garantizar los derechos fundamentales de los ciudadanos en el entorno digital. Aunque no fue diseñado específicamente para la IA, sus principios son plenamente aplicables a los sistemas que tratan datos personales mediante algoritmos automatizados.

Principios Fundamentales

El RGPD establece una serie de principios esenciales que deben cumplirse en todo tratamiento de datos personales, incluyendo aquellos realizados mediante sistemas de IA:

1. Licitud, lealtad y transparencia: el tratamiento debe tener una base legal clara y ser comprensible para el usuario.

2. Limitación de la finalidad: los datos tienen que ser recogidos con fines específicos, explícitos y legítimos.

3. Minimización de datos: solo se deben recopilar los datos necesarios para el propósito concreto.

4. Exactitud: se requiere que los datos sean exactos y estén actualizados.

5. Limitación del plazo de conservación: no deben conservarse más tiempo del necesario.

6. Integridad y confidencialidad: hay que tomar medidas de seguridad adecuadas para proteger los datos.

7. Responsabilidad proactiva: el responsable del tratamiento debe ser capaz de demostrar el cumplimiento de todos estos principios.

Desafíos de la IA

La aplicación del RGPD a la inteligencia artificial plantea una serie de dificultades:

• **Transparencia y explicabilidad:** Uno de los mayores retos es garantizar la transparencia en sistemas de IA complejos, especialmente aquellos que utilizan aprendizaje automático (machine learning) y aprendizaje profundo (deep learning). El RGPD exige que los interesados puedan entender cómo y por qué se toman ciertas decisiones automatizadas que les afectan (artículo 22). Sin embargo, muchos algoritmos operan como cajas negras, lo que dificulta ofrecer explicaciones claras.

• **Base legal del tratamiento**: Los sistemas de IA suelen alimentarse de grandes volúmenes de datos, y no siempre está claro qué base legal ampara este tratamiento: ¿consentimiento explícito?, ¿interés legítimo?, ¿ejecución de un contrato? Elegir y justificar la base legal adecuada es clave para garantizar la licitud del tratamiento.

• **Minimización y calidad de los datos:** La IA necesita grandes conjuntos de datos para aprender y generar resultados precisos. Esto puede entrar en conflicto con el principio de minimización. Además, los datos sesgados o incorrectos pueden llevar a decisiones discriminatorias, lo que vulnera el principio de exactitud y puede tener consecuencias legales y éticas.

• **Derechos de los interesados:** El RGPD otorga a los ciudadanos derechos como el acceso, rectificación, supresión de datos y oposición al tratamiento. En el contexto de la IA, garantizar el ejercicio efectivo de estos derechos requiere desarrollar mecanismos técnicos y organizativos adecuados, lo que no siempre es sencillo.

Buenas prácticas y enfoque ético

Para mitigar estos desafíos, las organizaciones que desarrollan o utilizan inteligencia artificial deben adoptar buenas prácticas de gobernanza de datos:

• **Evaluaciones de impacto de protección de datos:** son obligatorias en tratamientos automatizados de alto riesgo, como los que afectan significativamente a las personas. Permiten identificar riesgos y establecer medidas de mitigación.

• **Diseño ético y por defecto:** implementar desde el inicio principios de privacidad por diseño y por defecto.

• **Explicabilidad algorítmica:** es desarrollar sistemas de IA que sean comprensibles para los humanos, aunque se sacrifiquen niveles mínimos de precisión.

• **Supervisión humana:** garantizar que siempre haya una posibilidad real de intervención humana en decisiones automatizadas.

Hacia un futuro regulador integrado

La Unión Europea está avanzando hacia una regulación más específica para la IA, como el Reglamento de Inteligencia Artificial (RIA). Esta nueva normativa complementa al RGPD, estableciendo requisitos concretos para los sistemas de IA de alto riesgo, incluyendo obligaciones de transparencia, documentación y supervisión humana. Se espera que ambas normativas trabajen de manera conjunta para fomentar una IA ética, segura y respetuosa con los derechos fundamentales.

4. Transparencia y derechos del interesado

La expansión de la inteligencia artificial trae consigo grandes avances tecnológicos, pero también nuevos desafíos en materia de protección de datos personales. En este contexto, la transparencia y los derechos del interesado se posicionan como pilares fundamentales para garantizar que el uso de estas tecnologías respete la privacidad y la dignidad de las personas.

¿Qué implica la transparencia en la IA?

La transparencia es uno de los principios rectores del tratamiento de datos personales, especialmente cuando intervienen sistemas automatizados como los de inteligencia artificial. Se refiere a la obligación de informar de manera clara, accesible y comprensible a los usuarios sobre:

• Qué datos se recopilan.
• Con qué finalidad se usan.
• Quién los trata y durante cuánto tiempo.
• Si existen decisiones automatizadas y cuál es su lógica.
• Qué derechos tienen los interesados y cómo pueden ejercerlos.

En el contexto de la IA, este principio cobra una importancia aún mayor, dado que muchas veces los sistemas operan con algoritmos complejos, opacos o difíciles de interpretar para el ciudadano común. Esto puede generar desconfianza y sensación de indefensión.

Derechos del interesado frente a la IA

El RGPD europeo otorga a las personas una serie de derechos para que conserven el control sobre su información personal. Entre los más relevantes frente a los sistemas de IA, destacan:

• **Derecho a la información:** El interesado tiene derecho a saber, desde el primer momento, que sus datos están siendo recopilados y tratados. En el caso de la inteligencia artificial, debe informarse si hay decisiones automatizadas y cuál es la lógica utilizada.

• **Derecho de acceso**: Permite al interesado conocer qué datos suyos posee el responsable del tratamiento, cómo se están utilizando y si se están compartiendo con terceros.

• **Derecho a la rectificación y supresión**: Faculta corregir datos inexactos o a solicitar que se eliminen cuando ya no sean necesarios o se haya retirado el consentimiento para su uso.

• **Derecho a la oposición y limitación del tratamiento**: Permite oponerse al uso de sus datos en ciertos contextos, especialmente cuando se basan en intereses legítimos del responsable del tratamiento.

• **Derecho a no ser objeto de decisiones automatizadas**: Como ya se ha indicado anteriormente, el artículo 22 del Reglamento General de Protección de Datos establece que las personas tienen derecho a no ser objeto de decisiones basadas únicamente en tratamientos automatizados, incluida la elaboración de perfiles, que produzcan efectos jurídicos o les afecten significativamente. Además, tienen derecho a una intervención humana significativa, a expresar su punto de vista y a impugnar la decisión.

Desafíos actuales

A pesar del reconocimiento normativo de estos derechos, su aplicación práctica frente a sistemas de inteligencia artificial no siempre es sencilla. Algunos de los retos incluyen:

• Falta de comprensión de los algoritmos.

• Dificultad para explicar decisiones complejas generadas por modelos de aprendizaje automático.

• Limitaciones técnicas y económicas para implementar sistemas explicables y auditables.

Por ello, se habla cada vez más de la necesidad de una IA explicable (XAI, por sus siglas en inglés), que permita entender cómo y por qué una IA ha tomado una determinada decisión.

5. Decisiones automatizadas y perfilado

En el mundo digital actual, donde los algoritmos y modelos de inteligencia artificial toman un papel cada vez más protagónico, las decisiones automatizadas y el perfilado se han convertido en herramientas frecuentes en sectores como la banca, la publicidad, los recursos humanos, la

salud y muchos más. Estas prácticas, impulsadas por la inteligencia artificial, ofrecen eficiencia y personalización, pero también plantean importantes desafíos éticos y jurídicos, especialmente en lo que respecta a la protección de los datos personales.

¿Qué son las decisiones automatizadas?

Las decisiones automatizadas son aquellas que se toman sin intervención humana significativa, es decir, decisiones en las que un sistema informático –basado en algoritmos o inteligencia artificial– evalúa datos y produce un resultado que afecta a una persona. Algunos ejemplos incluyen:

• La aprobación o denegación de un crédito.
• La selección de candidatos en un proceso de contratación.
• La aplicación de precios personalizados en plataformas en línea.
• El bloqueo automático de cuentas por supuesta actividad sospechosa.

¿Y qué es el perfilado?

El perfilado (profiling) es un proceso automatizado que consiste en analizar datos personales para evaluar ciertos aspectos de una persona, como su comportamiento, intereses, preferencias, salud, rendimiento laboral, ubicación o situación económica.

Por ejemplo, una tienda en línea puede crear perfiles de usuarios para mostrar productos que podrían interesarles, o una aseguradora puede establecer primas diferenciadas en función del comportamiento del asegurado.

Marco legal: ¿qué dice la normativa?

El Reglamento General de Protección de Datos de la Unión Europea dedica especial atención a toda esta serie de prácticas.

En concreto, el artículo 22, ya referenciado anteriormente, establece lo siguiente:

"El interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos sobre él o le afecte significativamente de modo similar".

Este derecho se aplica cuando no hay intervención humana real y la decisión tiene un impacto relevante sobre la persona (como la denegación de un préstamo, una oferta de empleo o una sanción).

Excepciones permitidas

El RGPD permite las decisiones automatizadas solo si se cumple al menos una de estas condiciones:

1. Necesidad contractual: cuando sea necesario para la celebración o ejecución de un contrato.

2. Autorización legal: si está permitido por una norma legal que prevea medidas adecuadas para proteger los derechos del interesado.

3. Consentimiento explícito: si la persona ha dado su consentimiento informado y específico para ello.

Incluso en estos casos, deben aplicarse garantías adicionales, como el derecho a la intervención humana, a expresar el punto de vista del interesado y a impugnar la decisión.

Riesgos del perfilado y la automatización

Aunque estas tecnologías ofrecen ventajas en términos de eficiencia y personalización, también presentan riesgos importantes:

• **Falta de transparencia:** muchas personas no saben que están siendo perfiladas o cómo se toman las decisiones.

• **Discriminación algorítmica:** si los datos utilizados contienen sesgos, los algoritmos pueden replicarlos o agravarlos.

• **Efectos sociales y psicológicos:** las personas pueden verse afectadas por decisiones injustas o invasivas sin posibilidad de defenderse.

• **Pérdida de control sobre la propia información** y sobre cómo se utiliza para tomar decisiones que impactan la vida diaria.

Hacia una IA más justa y explicable

La solución no está en rechazar el uso de IA, sino en promover una inteligencia artificial transparente, explicable y centrada en las personas.

Esto implica:

• Diseñar sistemas que permitan entender por qué se tomó una decisión.

• Informar claramente a los usuarios cuando estén sujetos a decisiones automatizadas o perfilado.

• Garantizar la posibilidad de intervención humana y de reclamación.

• Aplicar medidas técnicas y organizativas para evitar sesgos y errores.

6. Evaluación de impacto y principios de proporcionalidad

El uso creciente de sistemas de inteligencia artificial presenta grandes oportunidades para la innovación, pero al mismo tiempo plantea importantes riesgos para los derechos fundamentales, especialmente en lo relativo a la protección de datos personales.

Para hacer frente a estos desafíos, el Reglamento General de Protección de Datos establece mecanismos clave como la Evaluación de Impacto en la Protección de Datos (EIPD) y principios rectores como el de proporcionalidad, que garantizan que la tecnología (incluida la inteligencia artificial) se utilice de manera justa, responsable y respetuosa con la dignidad humana.

¿Qué es una Evaluación de Impacto en la Protección de Datos (EIPD)?

La Evaluación de Impacto (Data Protection Impact Assessment, DPIA) es una herramienta preventiva que permite analizar y gestionar los riesgos que un tratamiento de datos puede generar sobre los derechos y libertades de las personas. Su principal objetivo es anticiparse a los efectos negativos que puedan ocasionarse y adoptar las medidas necesarias para mitigarlos antes de que el tratamiento comience.

Según el artículo 35 del RGPD, esta evaluación es obligatoria cuando el tratamiento puede entrañar un alto riesgo para los derechos y libertades de las personas, especialmente en los siguientes casos:

• Evaluación sistemática y exhaustiva de aspectos personales, incluida la elaboración de perfiles.

• Tratamiento a gran escala de categorías especiales de datos (por ejemplo, datos de salud).

• Vigilancia sistemática de zonas de acceso público.

En el ámbito de la inteligencia artificial, donde los algoritmos y modelos pueden tomar decisiones automatizadas o procesar grandes volúmenes de datos sensibles, la EIPD se convierte en una herramienta esencial de gobernanza ética y legal.

¿Qué debe contener una EIPD?

Una evaluación de impacto debe incluir al menos:

1. Descripción del tratamiento: qué datos se tratan, con qué fines, y cómo se lleva a cabo el tratamiento.

2. Análisis de la necesidad y la proporcionalidad del tratamiento en relación con sus fines.

3. Evaluación completa y correcta de los

riesgos para los derechos y libertades de los interesados.

4. Medidas para mitigar los riesgos, como cifrado, anonimización, limitación de acceso, etc.

Además, se recomienda la consulta con el delegado de protección de datos (DPO) y, en casos de alto riesgo residual, la consulta previa con la autoridad de control.

Principio de proporcionalidad

El principio de proporcionalidad en la protección de datos, junto con el de necesidad, es un criterio clave en la evaluación de cualquier tratamiento de datos, especialmente en el uso de modelos de inteligencia artificial. Este principio establece que:

• El tratamiento debe ser adecuado para alcanzar el fin perseguido.

• Debe ser necesario, es decir, no debe haber otro medio menos invasivo para lograrlo.

• Debe ser también proporcional, equilibrando los intereses del responsable con los derechos de los interesados.

En el contexto de la inteligencia artificial, este análisis es fundamental para evitar prácticas excesivas o invasivas, como el uso de reconocimiento facial en espacios públicos o el perfilado masivo sin consentimiento explícito.

Decisiones automatizadas y necesidad de control

Los sistemas de inteligencia artificial tienen la capacidad de tomar decisiones que afectan directamente a las personas (por ejemplo, en procesos de selección laboral, asignación de beneficios sociales o cálculo de primas de seguros).

En estos casos, una EIPD bien realizada y un análisis riguroso de proporcionalidad son esenciales para:

• Evaluar si la automatización es legítima y justificada.

• Detectar posibles sesgos o discriminaciones algorítmicas.

• Garantizar derechos como el de información, oposición, y revisión humana.

7. Riesgos éticos y supervisión algorítmica

La inteligencia artificial ya no es una promesa del futuro: hoy forma parte de decisiones que afectan a nuestra vida diaria, desde recomendaciones en plataformas digitales hasta evaluaciones de crédito o diagnósticos médicos. Pero con este avance tecnológico surgen también riesgos éticos que no deben subestimarse. La protección de datos personales y la supervisión algorítmica emergen como mecanismos clave para garantizar que la IA se desarrolle y aplique con responsabilidad y respeto por los derechos fundamentales.

Principales riesgos éticos de la IA

El uso de sistemas automatizados y algoritmos puede derivar en situaciones que comprometan valores esenciales como la justicia, la autonomía, la no discriminación o la privacidad. Entre los principales riesgos éticos destacan:

• **Discriminación algorítmica:** Los sistemas de IA pueden reproducir o incluso amplificar sesgos existentes si son entrenados con datos históricos desbalanceados. Esto puede llevar, por ejemplo, a decisiones discriminatorias en procesos de selección laboral, concesión de hipotecas o asignación de beneficios sociales.

• **Falta de transparencia y explicabilidad:** Muchos algoritmos y modelos de inteligencia artificial funcionan como cajas negras, lo que dificulta entender cómo se llega a una decisión. Esto vulnera el derecho a la información y la

posibilidad de cuestionar resultados que afectan directamente a las personas.

• **Pérdida de autonomía**: La dependencia de sistemas automatizados puede reducir la capacidad de las personas para tomar decisiones libres e informadas, especialmente cuando se enfrentan a algoritmos opacos o que se presentan como infalibles.

• **Vigilancia masiva y erosión de la privacidad**: El uso de tecnologías como el reconocimiento facial o el rastreo de comportamiento digital plantea serios dilemas sobre la proporcionalidad de estas prácticas y su impacto en la vida privada.

Protección de datos como salvaguarda ética

El Reglamento General de Protección de Datos de la Unión Europea no solo regula el tratamiento de información personal, sino que incorpora principios éticos fundamentales, ya señalados más arriba de este artículo, que son especialmente relevantes en el contexto de la IA. Estos principios, aunque sea una repetición, son:

• **Licitud, lealtad y transparencia:** las personas deben saber qué datos se recogen, para qué se usan y cómo se toman las decisiones automatizadas.

• **Minimización de datos:** solo deben tratarse los datos estrictamente necesarios.

• **Limitación del tratamiento automatizado:** el artículo 22 del RGPD reconoce el derecho a no ser objeto de decisiones basadas únicamente en el tratamiento automatizado sin intervención humana significativa.

Supervisión algorítmica

La supervisión algorítmica se refiere a la necesidad de establecer mecanismos de control humano, técnico y jurídico sobre los sistemas y modelos de inteligencia artificial para garantizar que operen de forma ética, justa y conforme a la legalidad.

Esta supervisión debe ser proactiva, no solo reactiva, e implica:

• **Auditorías de algoritmos:** evaluaciones periódicas para detectar errores, sesgos o resultados no deseados.

• **Explicabilidad de los modelos:** diseñar sistemas que puedan justificar sus decisiones de forma comprensible para los usuarios.

• **Participación interdisciplinar**: incluir en los equipos de desarrollo no solo a técnicos, sino también a juristas, filósofos, sociólogos y representantes de los usuarios.

• **Transparencia institucional:** las entidades públicas y privadas deben rendir cuentas sobre el uso de IA en sus procesos.

Marco europeo para una IA ética

El RIA refuerza esta visión al establecer obligaciones específicas para los sistemas de "alto riesgo", incluyendo:

• Evaluaciones de impacto.
• Requisitos de trazabilidad y documentación técnica.
• Supervisión humana obligatoria.
• Prohibición de usos que atenten contra los derechos fundamentales.

En conjunto con el RGPD, se está configurando en Europa un ecosistema normativo pionero, que busca equilibrar la innovación con la ética y la protección de las personas.

8. Reglamento de IA europeo

La inteligencia artificial ha pasado de ser un concepto futurista a una realidad cotidiana que influye en nuestras decisiones, trabajo, salud, educación y relaciones sociales. Frente a esta revolución tecnológica, la Unión Europea ha dado un paso histórico al aprobar el Reglamento europeo de Inteligencia Artificial (Artificial Intelligence Act), el primer marco legal del

mundo centrado exclusivamente en regular la IA. Su objetivo: fomentar la innovación sin renunciar a los derechos fundamentales, especialmente la protección de datos personales.

¿Qué es el Reglamento Europeo de IA?

El RIA es una propuesta legislativa presentada por la Comisión Europea en abril de 2021, cuya versión final ha sido objeto de acuerdos políticos y se espera su entrada en vigor en 2025. Su finalidad es crear un marco jurídico armonizado para el desarrollo, comercialización y uso de sistemas de IA en la Unión Europea.

A diferencia del Reglamento General de Protección de Datos, que regula el tratamiento de datos personales, el Reglamento de IA se enfoca en el comportamiento y riesgos de los sistemas de inteligencia artificial, independientemente de si tratan datos personales o no.

Enfoque basado en el riesgo

El Reglamento clasifica los sistemas de IA según su nivel de riesgo para los derechos fundamentales, estableciendo obligaciones:

1. Riesgo inaceptable (prohibidos):
- o Sistemas que manipulan conductas mediante técnicas subliminales.
- o Clasificación social según autoridades públicas (social scoring).
- o Reconocimiento facial en tiempo real en espacios públicos (con excepciones muy limitadas).

2. Riesgo alto (estrictamente regulados):
- o IA en contextos críticos como contratación laboral, educación, salud, justicia o gestión de fronteras.
- o Obligación de realizar evaluaciones de conformidad, garantizar transparencia y supervisión humana.

3. Riesgo limitado (obligaciones mínimas):
- o Sistemas o modelos que interactúan con humanos (es el ejemlo de los asistentes virtuales).
- o Se exige informar claramente que se trata de una IA.

4. Riesgo mínimo o nulo:
- o La mayoría de las aplicaciones, como filtros de spam o recomendación, están permitidas sin requisitos específicos.

9. Datos personales, RIA y RGDP

La protección de datos personales sigue siendo un aspecto central en la regulación de la IA. De hecho, el RIA no sustituye al RGPD, sino que lo complementa. Esto significa que cualquier sistema de IA que trate datos personales debe cumplir simultáneamente con:

• Las exigencias del RGPD (transparencia, minimización, consentimiento, etc.).

• Las obligaciones del Reglamento de IA (control del riesgo, supervisión humana, documentación técnica, etc.).

Esta interacción plantea varios puntos críticos:

• **Decisiones automatizadas y derechos del interesado**: Tanto el RGPD como el Reglamento de IA ponen límites a las decisiones tomadas únicamente por máquinas cuando afectan significativamente a las personas. Esto incluye el derecho a no ser objeto de decisiones automatizadas sin supervisión humana real y a recibir explicaciones comprensibles.

• **Evaluaciones de impacto conjuntas**: En muchos casos, será necesario realizar tanto una Evaluación de Impacto en Protección de Datos (EIPD) como una evaluación de conformidad del sistema de IA, especialmente en aplicaciones de alto riesgo. Esto obliga a coordinar equipos técnicos, legales y éticos desde la fase de diseño.

• **Transparencia algorítmica**: Ambos reglamentos refuerzan el principio de transparencia. No basta con usar IA de forma

eficaz: hay que explicar cómo funciona, qué datos utiliza, cómo se toman las decisiones y cuáles son los efectos para las personas.

Un enfoque ético, seguro y humano

La UE no sólo quiere liderar la innovación tecnológica, sino hacerlo de forma ética. Por eso, el Reglamento de IA se inspira en los valores europeos: dignidad humana, libertad, igualdad y Estado de Derecho.

Además, se promueve el concepto de "IA confiable", que implica:

- Supervisión humana significativa.
- Seguridad y solidez técnica.
- Respeto a la privacidad y protección de datos.
- No discriminación ni sesgos algorítmicos.
- Responsabilidad y trazabilidad.

10. Retos para el sector Público y Privado

La expansión de la inteligencia artificial está transformando la forma en que instituciones públicas y empresas privadas prestan servicios, toman decisiones y gestionan la información. Sin embargo, este avance conlleva importantes desafíos en términos de privacidad, transparencia y derechos fundamentales. La protección de datos personales se convierte, así, en un pilar esencial para garantizar una innovación tecnológica ética, legítima y centrada en las personas.

Un marco legal exigente pero necesario

Tanto el Reglamento General de Protección de Datos (RGPD) como la legislación europea sobre IA imponen a los actores públicos y privados nuevas responsabilidades en el diseño y uso de tecnologías inteligentes. Estos marcos no buscan frenar el desarrollo digital, sino garantizar que dicho desarrollo respete la dignidad humana y los derechos fundamentales.

Esto supone un cambio de paradigma: la tecnología ya no puede desarrollarse sin considerar desde el principio aspectos como el consentimiento, la transparencia, la minimización de datos, o la posibilidad de intervención humana en decisiones automatizadas.

Retos específicos para el sector Público

El sector público enfrenta una doble responsabilidad: debe promover la digitalización y al mismo tiempo ser garante del respeto a los derechos de la ciudadanía. Entre sus principales desafíos están:

- **Uso ético de sistemas automatizados en servicios públicos**: Administraciones que adoptan sistemas de IA para gestionar beneficios sociales, predicción de delitos o control migratorio deben asegurarse de que estas herramientas no generen discriminación ni violen derechos fundamentales.

- **Transparencia y rendición de cuentas**: Las decisiones adoptadas por algoritmos en la administración deben ser explicables, auditables y comprensibles para los ciudadanos. La falta de claridad mina la confianza pública.

- **Capacitación y gobernanza interna**: Muchos organismos públicos carecen de personal técnico y jurídico especializado en IA y protección de datos. Es clave fortalecer las capacidades internas y establecer comités éticos o unidades de supervisión algorítmica.

- **Evaluaciones de impacto en protección de datos**: El uso de IA por parte del sector público suele implicar tratamiento de datos sensibles o decisiones que afectan gravemente a las personas, por lo que es obligatorio realizar EIPD rigurosas.

Retos específicos para el sector Privado

El sector empresarial ha sido pionero en el uso de IA para mejorar procesos, reducir costes y personalizar servicios. Sin embargo, debe enfrentar retos importantes para garantizar un desarrollo responsable:

• **Cumplimiento normativo complejo**: Empresas que usan IA deben cumplir con múltiples obligaciones: respetar el RGPD, garantizar la seguridad de los datos, aplicar principios éticos y, en el caso de sistemas de alto riesgo, cumplir con los requisitos del Reglamento de IA.

• **Transparencia y confianza del consumidor**: La opacidad en el funcionamiento de algoritmos (por ejemplo, en precios dinámicos, recomendaciones o publicidad segmentada) puede generar desconfianza o rechazo por parte del consumidor. La explicabilidad y la gestión responsable del dato son claves para fortalecer la relación con los usuarios.

• **Discriminación algorítmica y reputación**: Los sistemas de inteligencia artificial mal entrenados pueden generar sesgos o decisiones injustas que dañen la imagen de la empresa, provoquen litigios o generen sanciones regulatorias. La auditoría de modelos y la diversidad de datos son estrategias clave para mitigar este riesgo.

• **Innovación ética como ventaja competitiva**: El mercado valora cada vez más las empresas que apuestan por una IA ética y responsable, que protege los datos y promueve un impacto social positivo. Convertir la ética en parte del modelo de negocio ya no es una opción, sino una ventaja competitiva.

Una oportunidad compartida

Tanto el sector público como el privado tienen la oportunidad y la responsabilidad de construir un ecosistema digital donde la inteligencia artificial no sólo sea eficiente, sino también justa, transparente y segura.

La clave está en fomentar una cultura de protección de datos desde el diseño, incorporando así la ética en todas las fases del desarrollo tecnológico, logrando garantizar mecanismos efectivos de supervisión y participación ciudadana.

11. Conclusiones

La inteligencia artificial no es una moda pasajera, sino una herramienta poderosa que ya está transformando el mundo. Comprender qué es, cómo funciona y hasta dónde puede llegar es esencial para aprovechar sus beneficios y enfrentar sus retos con responsabilidad. El futuro de la IA está en nuestras manos, y cómo decidamos usarla será clave para construir una sociedad más justa, eficiente y humana.

El RGPD representa un pilar esencial para garantizar que la inteligencia artificial se desarrolle y utilice en consonancia con los derechos de los ciudadanos europeos. Aunque su aplicación al campo de la IA presenta desafíos técnicos y conceptuales, también ofrece una base sólida para construir tecnologías responsables y centradas en el ser humano. El futuro de la IA en Europa dependerá, en gran medida, de la capacidad para armonizar innovación y protección de datos en un entorno cada vez más automatizado.

En una sociedad digital guiada por algoritmos, garantizar la transparencia y los derechos del interesado no es sólo una exigencia legal, sino una condición esencial para una IA ética y confiable. Las personas deben seguir siendo el centro del sistema, no meros objetos de análisis estadístico.

El reto está en diseñar tecnologías que no sólo sean eficientes, sino también justas, comprensibles y respetuosas con los derechos fundamentales. Porque el verdadero progreso no está en lo que la inteligencia artificial puede hacer, sino en cómo decidimos usarla.

Las decisiones automatizadas y el perfilado pueden ser herramientas útiles y eficientes, pero su uso debe equilibrarse con el respeto a los derechos fundamentales y a la autonomía de las personas. La protección de datos personales no es un freno a la innovación, sino una condición esencial para una transformación digital ética, justa y sostenible.

Porque en la era de los algoritmos, la protección de datos es también protección de la dignidad humana.

La inteligencia artificial tiene el potencial de transformar positivamente nuestras sociedades, pero solo si se desarrolla y utiliza dentro de un marco de ética, legalidad y respeto por los derechos fundamentales. La Evaluación de Impacto en Protección de Datos y los principios de proporcionalidad y necesidad no son obstáculos, sino herramientas indispensables para asegurar una IA transparente, responsable y centrada en el ser humano.

Estos principios evitan sanciones legales y fomentan la confianza en las nuevas tecnologías. Porque en la era de los algoritmos, la verdadera innovación no está solo en lo que la inteligencia artificial puede hacer, sino en cómo lo hace y a quién sirve.

La IAtiene un inmenso potencial transformador, pero sin un enfoque ético y sin mecanismos eficaces de supervisión algorítmica, sus beneficios pueden verse eclipsados por sus riesgos. La protección de datos personales no debe verse como un obstáculo, sino como una herramienta clave para diseñar una IA que respete la libertad, la igualdad y la dignidad humana.

En definitiva, el verdadero avance no radica solo en crear algoritmos más potentes, sino en construir tecnologías más justas, transparentes y humanas.

El RIA representa un hito en la regulación de las nuevas tecnologías, y su interacción con el RGPD refuerza la necesidad de un desarrollo digital centrado en las personas. La protección de datos no es un obstáculo, sino una garantía esencial de confianza y legitimidad.

Frente a un mundo cada vez más automatizado, Europa apuesta por una serie de modelos de inteligencia artificial que no solo sean inteligentes, sino también justos, transparentes y respetuosos con los derechos fundamentales. Porque el futuro tecnológico será sostenible solo si también es humano.

Reskilling y upskilling de los profesionales de RR. HH. para implantar con éxito la IA

Olga Herrero, experta en Formación.
Miembro del Consejo Asesor del Proyecto IA+Igual.

El hecho digital, en general, y la inteligencia artificial (IA), en particular, están transformando radicalmente el panorama empresarial. Los departamentos de Recursos Humanos no son la excepción: la IA está pasando de ser una tecnología emergente a convertirse en una herramienta fundamental que redefine cómo se gestionan las personas en las organizaciones.

"For us as Human Resources leaders, AI will change every process we touch. The way we source, assess, hire, train, develop, pay, and move people is all being informed by AI"
Josh Bersin (2019)

Según Randstad (2024), la IA aplica a todas las etapas del ciclo de vida del talento en las organizaciones (gráfico inferior).

Es evidente que una disrupción como la escrita requiere una adaptación de las competencias de los profesionales implicados en esta nueva realidad, generando una brecha entre las competencias disponibles en el mercado laboral y las requeridas por las organizaciones (skills gap). La responsabilidad de los equipos de Recursos Humanos no se limita a acompañar la implantación de las tecnologías emergentes, sino que conlleva el desarrollo de una mentalidad que permita aprovechar estas herramientas para crear valor estratégico. Así, las necesidades de aprendizaje de los miembros de estos equipos van mucho más allá de la adquisición o actualización de conocimientos técnicos.

La transformación del ecosistema digital no solo está reconfigurando los roles tradicionales de los profesionales que se ocupan de la gestión de personas, sino que también está creando nuevas posiciones especializadas. Los conceptos de reskilling y upskilling emergen como una respuesta estratégica fundamental para los profesionales. Ambos enfoques están centrados en mejorar las habilidades de los empleados, aunque presentan diferencias:

Título: *Ciclo de la vida del empleado*



Fuente: Informe de IA y RR. HH., Randstad (2024)

• El upskilling o perfeccionamiento se centra en mejorar las habilidades de que dispone un profesional para que pueda realizar mejor su trabajo o asumir mayores responsabilidades en su rol actual. Se orienta a expandir las habilidades existentes para mejorar el desempeño en el trabajo actual o asumir tareas más avanzadas.

• El reskilling o recualificación implica poner en juego un conjunto completamente nuevo de habilidades para adaptarse a un nuevo rol.

En el caso de tecnologías emergentes como la IA, la implementación de programas de capacitación no solo mejora la productividad, sino que también posiciona a las empresas para mantenerse a la vanguardia de la innovación.

Responder a la IA

Los profesionales del área necesitan adquirir u optimizar un conjunto específico de competencias clave que les permita aprovechar al máximo estas tecnologías emergentes y garantizar una gestión efectiva del talento.

En este perfeccionamiento profesional juegan un papel importante varias categorías de competencias:

• **Competencias técnicas:** comprensión de sistemas de IA -cómo funcionan los algoritmos, aprendizaje automático y los sistemas de procesamiento de lenguaje natural-, conocimiento de sus posibles aplicaciones para la gestión de personas y habilidades de análisis de datos necesarias para extraer insights a partir de la información de los empleados con la finalidad de mejorar la toma de decisiones. No se trata de que los profesionales se conviertan en expertos técnicos en IA, sino en facilitadores estratégicos que puedan aprovechar estas tecnologías para transformar la gestión del talento.

• **Competencias analíticas:** formulación de las preguntas correctas a la IA -interpretación correcta de las respuestas generadas por ésta, identificación de patrones significativos y su traducción en estrategias accionables-.

• **Competencias estratégicas:** visión del impacto digital en el negocio -pensamiento sistémico o la protección de datos de los profesionales, que contribuyen a la alineación de cualquier iniciativa de inteligencia artificial con los objetivos de la organización-.

La transformación digital no es una cuestión únicamente de tecnología, sino fundamentalmente de personas: mientras la IA asume tareas repetitivas y mecanizables, las habilidades exclusivamente humanas se vuelven más valiosas que nunca.

La implantación de la IA en los departamentos de gestión de personas implica un cambio de paradigma: los profesionales y los sistemas inteligentes forman un equipo integrado y, entonces, este modelo colaborativo aprovecha lo mejor de ambos mundos: por una parte, la IA aporta velocidad de procesamiento, análisis de grandes volúmenes de datos y detección de patrones; por otra parte, los profesionales contribuyen con sus competencias *blandas* exclusivas.

Por ello, pese a la automatización que supone la IA, o precisamente a causa de la misma, las competencias *blandas* se convierten en un diferenciador esencial para los profesionales de Recursos Humanos. Algunas de estas competencias prioritarias son:

• **Mentalidad de crecimiento digital:** combinación del aprendizaje de nuevas herramientas con la adopción de una actitud positiva frente al cambio tecnológico y disposición para experimentar y aprender de los errores.

Esta mentalidad es clave en el área de personas, puesto que la actitud de los miembros del equipo hacia la tecnología influirá en cómo percibe y adopta toda la organización las herramientas digitales.

• **Adaptabilidad:** flexibilidad para ajustarse a nuevos flujos de trabajo y herramientas digitales.

• **Pensamiento crítico:** habilidad para evaluar la información proporcionada por sistemas de IA, identificando sus sesgos y contextualizando sus recomendaciones.

• **Comunicación empática:** transmitir mensajes complejos sobre los cambios tecnológicos de manera clara y considerada.

• **Inteligencia emocional:** consiste en reconocer las emociones propias y ajenas, algo que es fundamental para gestionar las reticencias que la inteligencia artificial puede despertar en el equipo humano.

"Los humanos aportarán valor donde las máquinas no puedan. A medida que nos encontramos con más y más IA, la inteligencia real, la empatía real y el sentido común real serán escasos. Los nuevos trabajos se basarán en saber cómo trabajar con máquinas, pero también en estas cualidades humanas únicas"
Satya Nadella, CEO de Microsoft

• **Creatividad:** imaginar nuevas aplicaciones de la tecnología centradas en la persona, donde la IA complementa en lugar de reemplazar.

• **Juicio ético:** prever y valorar las implicaciones éticas de las iniciativas tecnológicas considerando, entre otros, aspectos como la privacidad de datos, los sesgos algorítmicos, la transparencia o la dependencia del criterio humano en las decisiones críticas.

Estas serán clave para crear sinergias humano-tecnológicas que potencien el valor de las herramientas tecnológicas y centrarnos en los aspectos más humanos y efectivos de RR. HH.

De manera complementaria a estas competencias tradicionales, que se actualizan y complementan para dar respuesta a lo digital, los Recursos Humanos requieren nuevos roles como consecuencia de la IA. Estos roles pueden ser incorporados a partir del mercado laboral, pero también pueden ser asumidos por profesionales que amplían su responsabilidad en sus cargos y se reciclan para conseguirlo.

Aquí juega un papel fundamental el reskilling para asumir roles emergentes en el departamento de Recursos Humanos como los siguientes:

• **HR Digital Transformation Lead:** responsable de liderar la implementación de tecnologías digitales en procesos de RH. HH.

• **People Analytics Manager:** profesional especializado en analizar datos para tomar decisiones sobre capital humano basadas en evidencias.

• **Talent Intelligence Specialist:** experto en utilizar IA para identificar, atraer talento y retener talento estratégico.

Todos los conocimientos, habilidades y actitudes mencionados son necesarios de manera transversal para la incorporación de las tecnologías emergentes y, en particular, de la IA, a la gestión del capital humano.

Este tema también se puede abordar contemplando las capacidades en función de los **procesos**:

1. Reclutamiento y selección: son procesos tan esenciales como costosos porque implican una importante inversión de tiempo en actividades de búsqueda y evaluación de candidatos. Las herramientas basadas en IA permiten agilizar estos procesos a través de su automatización. Aparecen herramientas como los chatbots de reclutamiento, para la preselección inicial y la resolución de dudas de los candidatos, o los sistemas de

seguimiento de candidatos, que permiten escanear grandes volúmenes de currículums y seleccionar los perfiles más adecuados según los requerimientos específicos de cada puesto. La IA puede, además, facilitar la realización de entrevistas automatizadas.

Para aprovechar al máximo estas herramientas, los profesionales de RH. HH. necesitan desarrollar su conocimiento de estos sistemas para crear procesos fluidos y atractivos para las personas e interpretar de manera informada los resultados que generan. Así, competencias como el diseño de experiencia del candidato, la evaluación crítica de algoritmos para detectar sus posibles sesgos o el análisis de datos son claves para su implantación exitosa. Según Mercer (2024), el 76 % de los profesionales de RH. HH. considera que la IA ha mejorado la calidad de las contrataciones y el 58 % afirma que ha reducido el tiempo de contratación.

Pero, de nuevo, es necesario resaltar que la automatización integrada en los procesos de selección no reemplaza el criterio humano, sino que lo potencia traspasando tiempo de dedicación anteriormente utilizado en tareas repetitivas a aspectos estratégicos de la misma.

2. Acogida u onboarding: las organizaciones con un proceso de acogida sólido mejoran la retención de nuevos empleados en un 82 % y la productividad en más de un 70 %. La IA ha impactado en este proceso desde la automatización de las tareas administrativas hasta los chatbots de asistencia que responden preguntas 24*7, los sistemas de recomendación de contenido personalizado según el perfil y rol o el análisis predictivo para identificar posibles desafíos de adaptación.

El análisis de datos para interpretar las métricas que miden la efectividad del proceso y el progreso del nuevo colaborador, el diseño de experiencias digitales de acogida o la gestión de plataformas de onboarding son algunas de las competencias necesarias para diseñar e implementar procesos de integración efectivos en entornos digitales. Todo ello sin olvidar que la conexión emocional es el elemento crucial para una acogida exitosa.

3. Formación y desarrollo del talento: gracias a plataformas de aprendizaje que integran IA, las empresas pueden ofrecer programas de capacitación personalizados, que ajustan el contenido a las habilidades y el ritmo de aprendizaje de cada empleado, además de proporcionar retroalimentación instantánea y sugerir cursos adicionales para optimizar el rendimiento. Las simulaciones y entrenamientos interactivos enriquecen aún más los procesos de aprendizaje corporativo.

La implementación de la IA a los procesos de formación y desarrollo profesional no es un proyecto tecnológico, sino fundamentalmente una transformación en la cultura de aprendizaje organizacional.

Por ello, además de los conocimientos técnicos necesarios (sobre Leaning Management Systems –LMS-, itinerarios y contenidos adaptativos, etc.), los profesionales de gestión de personas requieren otros elementos como un liderazgo comprometido, recursos accesibles, espacios y tiempo dedicados o el reconocimiento del valor del aprendizaje.

4. Evaluación del desempeño: la IA puede complementar el juicio humano proporcionando datos para conseguir decisiones más precisas y justas, ya que permite recopilar y analizar gran cantidad de datos en tiempo real, ofreciendo elementos objetivos para valorar las fortalezas y áreas de mejora de los empleados.

Su utilización para analizar datos en tiempo real sobre la productividad y las interacciones en equipo, identificar patrones en los comentarios de evaluaciones 360°, ofrecer retroalimentación continua o alertar sobre riesgos de bajo rendimiento requieren de nuevas competencias por parte de los equipos de RR. HH., centradas en la comprensión de los algoritmos y modelos de IA aplicados, la Interpretación y análisis estadístico de los datos y el pensamiento crítico

para valorar las recomendaciones automatizadas.

En este sentido, es importante reseñar que el uso de IA para la evaluación del desempeño es una de las aplicaciones que provocan mayores reticencias entre los empleados, quienes temen por la privacidad de sus datos y la objetividad del proceso. Por ello, adquiere una importancia fundamental el establecimiento de protocolos claros y transparentes en su uso; así como la potenciación de los aspectos más humanos de este proceso, como las conversaciones de desarrollo y el reconocimiento personalizado.

5. Fidelización del talento: por último, la IA también facilita la fidelización de talento a través del uso de **tableros de análisis predictivo**, que identifican señales y patrones que pueden indicar problemas como desmotivación o falta de compromiso, lo que facilita al área de personas intervenir de manera proactiva para mitigar la rotación.

Además, hace posible la creación de estrategias personalizadas para la retención, centradas en aquellos factores prioritarios para cada empleado (como, por ejemplo, la conciliación, las oportunidades de crecimiento profesional o el buen ambiente de trabajo).

planes de capacitación personalizados

Pese a haber realizado una presentación genérica de competencias necesarias para el equipo de Recursos Humanos, el diseño de planes de perfeccionamiento y recualificación efectivos requiere un enfoque personalizado que considere las características específicas de cada perfil profesional.

La personalización de estos planes exige identificar los diferentes roles existentes (reclutadores, especialistas en compensación, socios de negocio, etc.), definir el nivel de interacción de cada uno de ellos con cada herramienta de IA y establecer claramente qué nivel de competencia debe alcanzar cada perfil según su rol y responsabilidades. Identificar dónde se encuentra cada miembro del equipo de Recursos Humanos en la curva de adopción de la inteligencia artificial permite diseñar estrategias específicas para facilitar la transición digital.

Modalidades de implementación

La implementación efectiva de estrategias de upskilling y reskilling requiere un enfoque sistémico que integre evaluación de competencias, diseño de itinerarios formativos personalizados, metodologías de aprendizaje experiencial y mecanismos de transferencia al puesto de trabajo.

Algunas de las modalidades más efectivas para estos programas de capacitación son:

- **Microlearning:** aprendizaje en pequeñas dosis, fácil de integrar en el día a día laboral.

- **Aprendizaje entre pares:** aprovecha el conocimiento interno y fortalece la colaboración.

- **Comunidades de práctica:** aprendizaje continuo y colaborativo.

- **Proyectos aplicados:** experiencias prácticas con tecnologías emergentes o de colaboración interdisciplinaria con equipos de tecnología en situaciones profesionales reales.

- **Estudio de casos:** análisis de situaciones que permitan desarrollar un pensamiento crítico sobre las implicaciones éticas de la inteligencia artificial.

Indicadores de éxito

En cuanto a las métricas que permiten valorar o evaluar los programas de capacitación y reciclaje implementados, son, entre otras, las siguientes:

- **Tasa de adopción** o grado de uso diario de las tecnologías emergentes sobre las que se ha capacitado al equipo.

• **Reducción del tiempo medio invertido** en los procesos de mantenimiento o de mejora de la calidad.

• **Grado de satisfacción** de las personas implicadas en los procesos.

• **Equilibrio adecuado** entre la automatización a través de la tecnología y el contacto humano en el proceso.

• **Mecanismos de detección e intervención** cuando los algoritmos muestran resultados sesgados o inadecuados.

• **Iniciativas o mejoras propuestas** por usuarios sobre los sistemas de inteligencia artificial.

• **Impacto en negocio** en términos de mejoras en KPIs específicos del departamento (tiempo de contratación, fidelización, etc.).

Referencias bibliográficas

Bersin, J. (2019). *HR in the age of AI: lots of change ahead*. https://joshbersin.com/2019/07/hr-in-the-age-of-ai-lots-of-change-ahead/

Business Insider (2020). *El rol del humano en la era de la IA: esto es lo que opinan los grandes líderes.* https://www.businessinsider.es/tecnologia/rol-humano-era-inteligencia-artificial-hablan-lideres-668156

Maxnuk A. (2024). *Cómo la inteligencia artificial está transformando a los Recursos Humanos*. Mercer. https://www.mercer.com/es-pe/insights/people-strategy/future-of-work/analisis-de-como-la-ia-transforma-a-los-recursos-humanos/

Randstad Research (2024). *IA y Recursos Humanos. Hacia una forma más inteligente y humana de gestión del talento*. https://www.randstadresearch.es/ia-recursos-humanos/

Aspectos éticos del uso de la IA en Recursos Humanos

Javier Moscoso, doctor en Filosofía y profesor de Investigación del CSIC.
Miembro del Consejo Asesor del Proyecto IA+Igual.

El uso de la inteligencia artificial en Recursos Humanos (RR. HH.) está transformando los procesos de contratación, evaluación y gestión del talento. Sin embargo, esta adopción plantea importantes retos éticos relacionados con la transparencia, la equidad, la privacidad y la responsabilidad. Este es el marco para explorar brevemente los principales riesgos éticos asociados al uso de modelos de inteligencia artificial (IA) en el área de talento, con especial énfasis en el sesgo algorítmico, la protección de datos personales y la necesidad de una ética de los datos a lo largo del ciclo de vida de los modelos de IA.

Introducción

La inteligencia artificial (IA) se ha convertido en una herramienta clave en la transformación digital de los Recursos Humanos, permitiendo automatizar tareas, optimizar procesos y mejorar la toma de decisiones en la gestión del talento. Al mismo tiempo, el despliegue de estos modelos no está exento de controversias éticas. El uso de algoritmos para evaluar candidatos, predecir el rendimiento o gestionar las trayectorias profesionales plantea preocupaciones razonables sobre la imparcialidad, la privacidad, la transparencia y la responsabilidad en la toma de decisiones automatizadas.

Dilemas éticos

Los principales dilemas éticos relacionados con el uso de la inteligencia artificial en el ámbito de los Recursos Humanos son los siguientes:

1. Sesgo algorítmico y discriminación: uno de los principales riesgos éticos del uso de la IA en gestión de personas es la reproducción o amplificación de los sesgos existentes en los datos de entrenamiento. Los algoritmos aprenden patrones de diferentes fuentes y datos, que pueden estar marcados por desigualdades estructurales. Esto puede dar lugar a decisiones discriminatorias hacia determinados grupos por motivos de género, etnia, edad o procedencia.

Ejemplos relevantes:

• *Amazon:* la empresa desarrolló un sistema de IA para evaluar currículos que penalizaba a las candidatas, ya que se entrenaba con datos históricos predominantemente masculinos (Dastin, 2018).

• *HireVue:* su sistema automatizado de entrevistas en vídeo utilizaba el análisis facial y de voz para puntuar a los candidatos, lo que suscitó dudas sobre su validez científica y posibles sesgos contra personas con discapacidad o dialectos diferentes (Harwell, 2019).

• *LinkedIn:* una investigación demostró que sus algoritmos de recomendación favorecían los perfiles masculinos para puestos ejecutivos, incluso cuando las mujeres tenían cualificaciones equivalentes (Cowgill, Dell'Acqua, & Deng, 2021).

Estos ejemplos, a los que podrían añadirse muchos otros, ponen de manifiesto la necesidad de incorporar criterios de equidad y auditabilidad

en el diseño y la evaluación de los sistemas de IA aplicados a los RR. HH.

El sesgo en la IA puede surgir no sólo de los datos de entrenamiento, sino también de las decisiones tomadas durante el desarrollo y la implantación del modelo. Por ejemplo, las variables indirectas que parecen neutrales, como los códigos postales o los nombres de los colegios, pueden servir como indicadores indirectos de raza o estatus socioeconómico, perpetuando la discriminación incluso cuando se excluyen los atributos sensibles. Este fenómeno, conocido como *discriminación indirecta*, es especialmente insidioso en contextos de contratación en los que la interpretabilidad del modelo es baja (Barocas y Selbst, 2016).

Además, la falta de diversidad demográfica entre los equipos de desarrollo puede limitar la concienciación sobre posibles sesgos y dar lugar a sistemas que pasen por alto las experiencias de los grupos infrarrepresentados. Por lo tanto, el diseño inclusivo y los procesos de revisión interdisciplinar son esenciales para evitar resultados excluyentes. También debe tenerse en cuenta la interseccionalidad, ya que las personas que pertenecen a múltiples grupos marginados (por ejemplo, mujeres de color, candidatos LGBTQ+ o de edad avanzada) son especialmente vulnerables a los prejuicios agravados (Crenshaw, 1991).

Otra cuestión crítica es el *efecto de retroalimentación*. Cuando los sistemas de IA se despliegan y comienzan a influir en los resultados del mundo real -como quién es contratado-, los datos de entrenamiento futuros se ven moldeados por estas decisiones algorítmicas, lo que puede reforzar los patrones de desigualdad existentes. Sin una supervisión y recalibración continuas, estos bucles pueden consolidar la desventaja estructural a lo largo del tiempo (Eubanks, 2018).

Para hacer frente a estos retos, las organizaciones deben ir más allá de las métricas básicas de imparcialidad y aplicar definiciones de imparcialidad específicas para cada contexto, como la *igualdad de oportunidades* o la *paridad demográfica*, dependiendo del caso de uso del área de talento. Las auditorías periódicas, las consultas a las partes interesadas y la integración de técnicas de aprendizaje que tengan en cuenta la imparcialidad son necesarias para garantizar un trato equitativo en todos los grupos demográficos.

2. Privacidad y protección de datos: el uso de la inteligencia artificial en Recursos Humanos implica inevitablemente el tratamiento intensivo de datos personales, lo que plantea importantes retos éticos y jurídicos relacionados con la privacidad. Las herramientas de IA requieren grandes volúmenes de información, desde datos tradicionales hasta datos biométricos o de comportamiento.

Uno de los principales riesgos es la *recopilación desproporcionada de datos,* que puede vulnerar el principio de minimización de datos establecido en el Reglamento General de Protección de Datos (RGPD) de la Unión Europea (Parlamento Europeo, 2016).

Además, la falta de transparencia y la opacidad de los sistemas algorítmicos dificultan el ejercicio de derechos fundamentales como el acceso, la rectificación o la supresión de los datos.

La seguridad de los datos es también una de las principales preocupaciones, debido al riesgo de ciberataques o usos no autorizados.

El uso de *fuentes de datos alternativas*, como las redes sociales o las grabaciones de entrevistas, plantea retos adicionales, ya que los candidatos a menudo no son conscientes de que se están tratando sus datos, lo que vulnera los principios básicos de consentimiento y autonomía.

Ejemplos de infracciones:

• *IBM:* utilizó bases de datos de imágenes sin consentimiento para entrenar modelos de reconocimiento facial (Vincent, 2020).

• *Xref:* retuvo evaluaciones de empleo sin notificación clara ni consentimiento, lo que suscitó preocupación por el almacenamiento indebido de datos sensibles (TechCrunch, 2022).

3. Consentimiento informado y autonomía de los datos: el consentimiento debe ser libre, específico, informado e inequívoco (Parlamento Europeo, 2016). En la práctica, sin embargo, los candidatos no suelen comprender el alcance del tratamiento de datos por parte de los sistemas de inteligencia artificial.

Casos como el de HireVue (véase más arriba) ilustran cómo el consentimiento formal no garantiza un consentimiento éticamente válido, ya que los participantes no son conscientes de que sus expresiones faciales y su tono de voz están siendo evaluados por algoritmos. Además, el consentimiento suele estar condicionado por la necesidad de acceder a una oportunidad de empleo, lo que plantea dudas sobre su libertad real (Binns, 2018).

La recogida de datos de las redes sociales sin previo aviso también constituye una violación de la autonomía informativa. Por lo tanto, es esencial desarrollar mecanismos de transparencia efectiva, supervisión humana y revocación voluntaria del consentimiento sin consecuencias adversas para las personas.

4. Transparencia y explicabilidad algorítmicas: la transparencia en los sistemas de IA aplicados a los Recursos Humanos es un principio fundamental para garantizar una toma de decisiones justa y responsable. La opacidad de los modelos algorítmicos, especialmente los basados en el aprendizaje profundo, dificulta la comprensión de cómo y por qué se toman determinadas decisiones.

Retos comunes*:*

• **Modelos de caja negra***:* muchos sistemas de IA funcionan como "cajas negras", sin ofrecer explicaciones claras sobre los criterios utilizados para puntuar o rechazar candidatos.

• **Falta de documentación***:* la ausencia de registros detallados sobre el diseño, la formación y el uso de los modelos dificulta la trazabilidad y las auditorías independientes.

• **Lenguaje técnico inaccesible***:* incluso cuando se dan explicaciones, suelen estar redactadas en términos muy técnicos, incomprensibles para los usuarios afectados.

Una mayor transparencia no sólo aumenta la confianza de los usuarios, sino que también permite detectar sesgos, errores y violaciones de derechos. También refuerza la responsabilidad organizativa a la hora de desplegar sistemas de IA en RR. HH.

5. Ética de los datos en el ciclo de vida de la IA: debe integrarse en todo el ciclo de vida de los sistemas de IA utilizados en RR. HH.:

• **Diseño del sistema:** la definición del problema debe abordarse desde una perspectiva pluralista e inclusiva. Los criterios para evaluar el talento o el rendimiento deben reflejar no solo la lógica empresarial, sino también valores de equidad e inclusión (Binns et al., 2018).

• **Recogida y conservación de datos:** es necesario aplicar los principios de proporcionalidad y calidad de los datos, evitando variables irrelevantes o discriminatorias. La revisión crítica de los datos históricos es esencial para evitar sesgos heredados (Mehrabi et al., 2021).

• **Formación y validación:** hacen falta técnicas de desviación y métricas éticas, más allá de las puntuaciones de precisión tradicionales.

Los modelos justos no son necesariamente los más eficientes en términos convencionales (Barocas, Hardt y Narayanan, 2019).

• **Aplicación:** garantizar una supervisión humana significativa. Ninguna decisión importante debe automatizarse por completo sin la posibilidad de revisión humana.

• **Evaluación continua:** los modelos deben auditarse periódicamente para evaluar los impactos imprevistos y corregir los efectos discriminatorios emergentes (Raji et al., 2020).

• **Documentación y explicabilidad:** una transparencia documental es esencial -con normas como Model Cards (Mitchell et al., 2019)

• **Datasheets for Datasets** (Gebru et al., 2021)- para comprender las características, limitaciones y supuestos de cada sistema.

Conclusión

El uso de la inteligencia en Recursos Humanos ofrece importantes ventajas, pero también riesgos éticos que no deben pasarse por alto. La discriminación algorítmica, las violaciones de la privacidad y la opacidad en la toma de decisiones son retos urgentes que requieren respuestas normativas y éticas integrales.

Es muy conveniente que las organizaciones adopten principios de ética de datos en todas las etapas del ciclo de vida de la inteligencia artificial, promoviendo la equidad, la transparencia, el consentimiento informado y la responsabilidad.

Solo a través de estas prácticas se podrá beneficiar la transformación digital de la gestión del talento a las personas de forma justa y sostenible.

Referencias bibliográficas

Barocas, S., y Selbst, A. D. (2016). *El impacto dispar de Big Data*. California Law Review, 104(3), 671-732.

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning*. http://fairmlbook.org

Binns, R. (2018). Equidad en el aprendizaje automático: Lecciones de la filosofía política. *Actas de la Conferencia 2018 sobre imparcialidad, rendición de cuentas y transparencia*, 149-159.

Bodie, M. T., Cherry, M. A., McCormick, M. L., & Tang, J. (2020). The Law and Policy of People Analytics. *University of Colorado Law Review*, 91(4), 961-1030.

Cowgill, B., Dell'Acqua, F., & Deng, S. (2021). ¿Programadores sesgados? Or Biased Data? A Field Experiment in Operationalizing AI Ethics. En *Proceedings of the 21st ACM Conference on Economics and Computation (pp. 679-681). Documento de investigación de la Columbia Business School*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3615404

Dastin, J. (2018). Amazon desecha herramienta secreta de reclutamiento de IA que mostró sesgo contra las mujeres. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Parlamento Europeo. (2016). Reglamento (UE) 2016/679 (Reglamento general de protección de datos).

Eubanks, V. (2018). *Automatizando la desigualdad: Cómo las herramientas de alta tecnología perfilan, vigilan y castigan a los pobres*. St. Martin's Press.

Floridi, L., & Cowls, J. (2019). Un marco unificado de cinco principios para la IA en la sociedad. *Harvard Data Science Review*, 1(1).

Gebru, T., et al. (2021). Hojas de datos para conjuntos de datos. *Communications of the ACM*, 64(12), 86-92.

Harwell, D. (2019). Un algoritmo de escaneo facial decide cada vez más si mereces el trabajo. *The Washington Post.* 6 de noviembre, consultado el 21 de abril de 2025 https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019). *Mejorar la imparcialidad en los sistemas de aprendizaje automático: Qué necesitan los profesionales de la industria?* CHI 2019.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6), 1-35.

Mitchell, M., et al. (2019). Tarjetas modelo para informes modelo. *Actas de la Conferencia sobre imparcialidad, rendición de cuentas y transparencia*, 220-229.

Raji, I. D., et al. (2020). Closing the AI Accountability Gap. *Actas de la Conferencia 2020 sobre imparcialidad, responsabilidad y transparencia*, 33-44.

TechCrunch. (2022). *La plataforma australiana Xref, bajo lupa por la gestión de los datos de los candidatos*.

Vincent, J. (2020). IBM usó fotos de Flickr para un proyecto de reconocimiento facial sin consentimiento. *The Verge*.

Wachter, S., Mittelstadt, B., y Floridi, L. (2017). Por qué no existe un derecho a la explicación de la toma de decisiones automatizada en el Reglamento General de Protección de Datos. *Derecho internacional de privacidad de datos*, 7(2), 76-99.

Zarsky, T. Z. (2016). El problema con las decisiones algorítmicas: Una hoja de ruta analítica para examinar la eficiencia y la equidad en la toma de decisiones automatizada y opaca. *Science, Technology, & Human Values,* 41(1), 118-132.

Zuboff, S. (2019). *La era del capitalismo de la vigilancia*. PublicAffairs.

Cultura del dato en RR. HH.: de los indicadores de desempeño a los modelos de IA

Ana Valera, experta en People Analytics.
Miembro del Consejo Asesor del Proyecto IA+Igual.

En los últimos años, la gestión de personas ha experimentado un cambio tan profundo como inevitable. Más allá del protagonismo que han cobrado las tecnologías digitales, lo que realmente está transformándose es la forma en la que pensamos, analizamos y decidimos dentro de las organizaciones. La transición desde una lógica basada en la intuición hacia una fundamentada en la evidencia, y desde un enfoque reactivo hacia una mirada anticipatoria, ha convertido a la cultura del dato en un eje estructural de los departamentos de Recursos Humanos (RR. HH.). No se trata de una moda ni de una herramienta más: hablamos de un cambio de paradigma que interpela tanto a las prácticas como a los principios que las sustentan.

Esta es una reflexión sobre qué significa realmente contar con una cultura del dato en el ámbito de la gestión de personas, comprender cómo ha evolucionado a lo largo del tiempo y por qué constituye un pilar indispensable para que la inteligencia artificial (IA) pueda aportar valor real y duradero.

De las métricas operativas a la analítica estratégica

Durante décadas, los departamentos de Recursos Humanos han trabajado con un repertorio relativamente limitado de indicadores: la rotación, el absentismo, la duración media de los procesos de selección o los costes asociados a la formación. Si bien estas métricas han sido útiles para controlar ciertos aspectos operativos, su capacidad para generar conocimiento estratégico ha sido reducida. En muchos casos, nos decían qué había pasado, pero no nos ayudaban a entender por qué ocurrió ni qué podíamos hacer al respecto.

Autores como Rasmussen y Ulrich (2015) insisten en que el verdadero aporte del área de talento empieza cuando deja de limitarse a recopilar y presentar datos, y asume un papel activo como intérprete y generador de conocimiento. En esta misma línea, el modelo de madurez analítica desarrollado por Bersin (2016) describe una evolución que va desde el uso básico y descriptivo de los datos hacia un estadio en el que la analítica predictiva y prescriptiva se convierte en parte natural del proceso de toma de decisiones.

Pese a estos avances conceptuales, la realidad práctica muestra una adopción aún desigual. Según datos de Deloitte (2021), más del 60 % de las organizaciones sigue anclada en los niveles más básicos de esta escala de madurez, lo que pone de manifiesto la necesidad de impulsar no solo la tecnología, sino también las capacidades y la mentalidad que la acompañan.

Evolución de la cultura del dato

El recorrido que ha seguido la cultura del dato en RR. HH. refleja una transición paulatina, no exenta de tensiones.

En un primer momento, el uso de datos respondía a necesidades eminentemente administrativas: gestionar contratos, registrar ausencias, calcular nóminas. Estos datos solían almacenarse en papel o, con suerte, en hojas de Excel que cada

responsable actualizaba a su manera. Eran tiempos en los que el dato existía, sí, pero no hablaba.

La implantación de herramientas como Excel supuso una primera forma de sistematización, aunque limitada por la falta de estandarización, trazabilidad o posibilidad de análisis cruzado. El verdadero punto de inflexión llegó con los Sistemas de Gestión de Recursos Humanos (Human Resource Management Systems, sigla en inglés HRMS), que facilitaron la centralización de la información y la generación de indicadores clave de manera más fiable. No obstante, en muchos casos, el uso seguía siendo reactivo: se generaban informes cuando alguien los pedía, se reportaban cifras para justificar decisiones ya tomadas.

La transformación más significativa se ha producido en los últimos años, con la llegada de herramientas de análisis avanzado y, más recientemente, con el uso de modelos de IA. Algunas organizaciones han comenzado a utilizar sus datos no solo para describir lo ocurrido, sino para anticipar escenarios, explicar fenómenos complejos y automatizar respuestas. Hablamos, por ejemplo, de predecir qué perfiles tienen mayor riesgo de abandonar la empresa, de identificar qué características comparten los empleados con mejor desempeño, o de sugerir planes de desarrollo personalizados en función de trayectorias previas.

Todo esto no es, en el fondo, una revolución tecnológica. Es, sobre todo, una evolución cultural: pasar de registrar a comprender, de gestionar procesos a gestionar personas en toda su complejidad.

Papel de RR. HH. en la gobernanza del dato

Este cambio también ha venido acompañado por una creciente conciencia sobre la necesidad de establecer marcos sólidos de Gobernanza del dato. Hasta hace poco, esta responsabilidad recaía casi exclusivamente en las áreas técnicas. Sin embargo, la naturaleza de la información que maneja Recursos Humanos —profundamente sensible y directamente vinculada a la experiencia y los derechos de las personas— exige que este departamento asuma un rol activo y propositivo.

Los datos de RR. HH. no se limitan a cifras objetivas: también recogen percepciones, emociones y patrones de conducta. Utilizarlos sin el debido cuidado puede generar consecuencias éticas, legales y reputacionales difíciles de revertir. Por eso, además de garantizar la seguridad y el cumplimiento normativo, RR. HH. debe liderar el diseño de políticas que aseguren un uso transparente, equitativo y explicable de los algoritmos.

Floridi et al. (2018) abogan por una IA centrada en principios éticos como la justicia, la no discriminación y la comprensibilidad de los procesos automatizados. RR. HH., por su proximidad al impacto real de las decisiones, está especialmente capacitado para asumir ese liderazgo. Además, puede actuar como traductor entre el lenguaje técnico de los datos y la realidad vivida por los equipos humanos.

También tiene una tarea clave como facilitador del cambio cultural. Impulsar la alfabetización en datos, fomentar la curiosidad analítica y promover la toma de decisiones fundamentada son responsabilidades que deben integrarse en el día a día del área.

Una mirada final

Construir una cultura del dato sólida en capital humano no se consigue únicamente implementando nuevas tecnologías. Es una tarea mucho más profunda, que implica redefinir creencias, transformar rutinas y asumir un compromiso firme con una gestión basada en la comprensión y el respeto.

Hoy, más que nunca, el área de personas tiene la oportunidad -y la responsabilidad- de guiar el uso responsable de los datos en las organizaciones.

No solo por su capacidad para interpretarlos, sino por lo que representa: una mirada humanista, centrada en el valor de cada persona.

Como recuerda Minbaeva (2018), el valor de la analítica en el ámbito de los recursos humanos no está en su sofisticación matemática, sino en su capacidad para generar confianza, legitimidad y sentido. Una cultura del dato no es un fin en sí mismo, es más bien una herramienta para tomar decisiones más justas, más inteligentes y, sobre todo, más humanas.

Referencias bibliográficas

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal*, 26(1), 1–11. https://doi.org/10.1111/1748-8583.12090

Bersin, J. (2016). *Talent Analytics Maturity Model*. Bersin by Deloitte.

Deloitte. (2021). *Global Human Capital Trends*. Deloitte Insights. https://www2.deloitte.com

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society. *Minds and Machines*, 28(4), 689–707. *https://doi.org/10.1007/s11023-018-9482-5*

Levenson, A. (2018). *Strategic Analytics: Advancing Strategy Execution and Organizational Effectiveness*. Berrett-Koehler Publishers.

Minbaeva, D. B. (2018). Building credible human capital analytics for organizational competitive advantage. *Human Resource Management*, 57(3), 701–713. https://doi.org/10.1002/hrm.21858

Parlamento Europeo. (2024). *Reglamento de Inteligencia Artificial*. Diario Oficial de la Unión Europea.

Rasmussen, T., & Ulrich, D. (2015). Learning from practice: How HR analytics avoids being a management fad. *Organizational Dynamics*, 44(3), 236–242. https://doi.org/10.1016/j.orgdyn.2015.05.008

Zicari, R. V., Brodersen, C., Brusseau, J., & Sætra, H. S. (2021). *A European Proposal for Trustworthy AI: A Framework. Springer*. https://doi.org/10.1007/978-3-030-81907-1

Cómo afecta la inteligencia artificial a los procesos del área de talento

Carlos Cabañas, vice president of Sales and Marketing en denario.
Embajador del Proyecto IA+Igual.

"Anything a typical person can do in less than 1 second of thought, we can probably now (or soon) automate"
Andrew Ng. (2017)

Si hay un hecho indiscutible es que la inteligencia artificial (IA) está impactando ya de forma significativa en todos nuestros ámbitos de actividad; obviamente, la gestión de personas no es una excepción, sino más bien todo lo contrario.

¿De qué manera?

Para plantear una reflexión ordenada, podríamos pensar en identificar el tipo de actividades que se pueden ver especialmente afectadas en función de su naturaleza, y también en contemplar aquellas facetas que, sin ser a priori propias del área de gestión de personas (o de Recursos Humanos, llámelo usted como quiera pero sin perder de vista su aportación de valor), sí que terminan suponiendo un cambio significativo en lo que al desarrollo de la carrera profesional o actividad de los profesionales se refiere.

En relación con las actividades que pueden verse afectadas por el uso de la IA, comencemos identificando la naturaleza de las tareas en las que, a priori, la IA resulta especialmente efectiva.

Por un lado, podríamos hablar de aquellas en la que las que los humanos somos buenos, pero lentos, con un capítulo especial para todas esas que odiamos hacer o que son aburridamente repetitivas. Por ejemplo, ¿quién no sabe distinguir la imagen de un perro de la de un elefante?; pero, ¿qué pasa cuando hay que distinguir 10.000 imágenes de perros de 10.000 de elefantes, todas tiradas encima de la mesa? O si se trata de resumir un texto, seguro que todos, de mejor o peor manera, podemos hacerlo, sí; pero ¿y si tenemos que resumir una gran cantidad de documentos? Lo mismo pasa con temas aparentemente triviales como el reconocimiento de voz: podemos diferenciar las voces, pero el volumen y la repetición nos matan…

Por otro lado, podríamos también identificar tareas en las que los humanos, directamente, somos malos, tenemos sesgos difíciles de evitar, o no tenemos capacidad para afrontarlas. Por ejemplo, aunque a priori podemos identificar patrones y relaciones entre datos con un número de dimensiones limitado, no ocurre lo mismo cuando ese número aumenta, o cuando pueden establecerse relaciones no lineales. Incluso en escenarios sencillos, hay problemas que no llegamos a solventar eficientemente sin ayuda cuando el tiempo es un factor determinante.

Con esta sencilla clasificación, seguro que ya nos vienen a la cabeza algunas actividades propias del terreno de la gestión de personas en las que el uso de inteligencia artificial puede tener (-de hecho, ya lo tiene)- un impacto considerable. Algunos casos de uso, por

diferentes áreas de actividad, son los siguientes:

- **Selección:** automatización del filtrado inicial de CV, identificación y elección de perfiles ideales, búsqueda semántica en los documentos, entrevistas de evaluación con agentes virtuales, generación y publicación de ofertas de empleo atractivas, identificación de etapas con áreas de mejora en los procesos, generación de contratos, automatización de las diversas fases de comunicación con los candidatos, etc.

- **Acogida:** creación de materiales de onboarding, automatización de procesos burocráticos (alta en sistemas informáticos, asignación de equipos, cierre de agendas para reuniones con responsables de áreas, seguimiento…) y volcado automático a sistemas desde la fase de selección, son ejemplos de todo lo que se puede hacer con inteligencia artificial.

- **Formación:** generación de contenidos formativos (desde el índice hasta los contenidos elaborados propiamente dichos, enriquecidos gracias a las capacidades multimedia que ofrecen los nuevos sistemas de Inteligencia Artificial Generativa multimodales), asignación de itinerarios formativos personalizados e identificación de necesidades de formación, son solo algunas de las tareas automatizables.

- **Carrera profesional:** definición de perfiles de éxito, identificación de alto potencial, selección de candidatos ideales para un puesto o para un plan de sucesión, inventario de habilidades para facilitar la gestión del Smart Workforce Planning (planificación inteligente de la mano de obra), análisis y optimización de la productividad, etc.

- **Bienestar:** análisis avanzado del Plan de Bienestar, identificación de medidas exitosas (actuales y de previsión futura), optimización de inversiones, por citar algunos ejemplos.

- **Clima Laboral:** análisis de encuestas (contemplando tanto preguntas cuantitativas como cualitativas), identificación de patrones y de áreas de mejora y sugerencias para la creación de nuevas preguntas, entre otras tareas.

- **Transversal:** asistentes cognitivos (tipo ChatGPT enfocado a ofrecer respuestas y ayuda respecto a cuestiones propias del área de RR. HH.), por ejemplo.

Como puede comprobarse, todo un mundo de posibilidades, al que además habría que sumar, también, con alcance transversal, la capacidad que la IA ofrece para automatizar todas esas tareas rutinarias que habitualmente consumen una gran parte del tiempo de nuestros equipos como, por ejemplo, extraer datos de un sistema, cambiar el formato y pasarlo a otro… ¿Quién no se ha peleado con libros de Excel que vienen y van?

¿En qué se traduce este potencial impacto?

En primer lugar, sin duda, la tecnología (y en concreto, la IA) se configura como nuestra gran aliada para liberar tiempo de valor. Seremos capaces de dedicar nuestro tiempo a esas tareas de valor añadido a las que ahora simplemente no llegamos, porque nos lo quitan todas esas actividades rutinarias (pero imprescindibles) que nos persiguen cada día.

En segundo lugar, podría producirse un efecto colateral negativo, fruto de una deshumanización potencial provocada por el uso extensivo de la tecnología. Y aquí sí debemos tener cuidado. Dos recomendaciones que debemos tener siempre presentes: hay que prestar especial atención a los sesgos (que pueden estar provocados por los datos de origen con los que se entrenan los algoritmos, o por los algoritmos en sí); no podemos perder nunca de vista que la tecnología es una herramienta, pero que la responsabilidad última tras la toma de una decisión siempre recae en

las personas, nunca en las máquinas. La regulación europea sobre este segundo aspecto es especialmente vigilante e incluso categoriza el empleo de la IA en el ámbito de Recursos Humanos como de alto riesgo (de ahí que el apartado normativo regulatorio sea otra área de impacto no desdeñable).

Y, por último, hay otro potencial impacto del que quizá no se hable mucho, pero que me parece especialmente preocupante de cara al medio plazo, que tiene que ver con lo que enunciaba al principio: las implicaciones que, sin estar a priori ligadas a actividades propias del área de recursos humanos, sí que pueden afectar de forma significativa al desarrollo de la carrera profesional de las personas.

Me estoy refiriendo al mundo de juniors, de los perfiles con poca o nula experiencia, que dan los primeros pasos en su carrera. Hasta ahora, queriéndolo o no, esas personas comenzaban su andadura ocupándose de todas esas tareas rutinarias y repetitivas que recaían en "los nuevos" de forma natural. ¿Por qué? Porque no resultaban a priori complejas, bajo la supervisión de perfiles más senior, y porque también eran más baratas en manos de esos perfiles.

Lo malo de esas tareas es el aburrimiento que generaban a las personas jóvenes, la repetición, la sensación de no aportar valor, el hartazgo en muchas ocasiones. Pero resultan, a la vez, toda una escuela de aprendizaje. Quizá esos juniors no empezaran a trabajar con temas críticos, pero toda esa marabunta de documentos a los que se enfrentaban, todos esos cambios de formatos, esas columnas en tablas que van y que vienen, ese bucear en busca de datos y más datos, acababan impregnando a esos perfiles de conocimiento de negocio. Sentaba las bases para empezar a crecer.

¿Y ahora? Si todas esas tareas las van a hacer las máquinas, ¿cómo van a empezar a empaparse del negocio esos perfiles?, ¿cómo van a diseñarse las carreras?, ¿cómo se llegará a ese momento de aportar valor?

Sin duda, un montón de interrogantes a los que debemos dar respuesta de forma creativa y, por qué no, apoyándonos en la IA. Estamos viviendo un momento apasionante.

Normativa laboral española específica en torno a la inteligencia artificial

Estela Martín, Abogada y DirCom de Sincro.
Embajadora del Proyecto IA+Igual.

1. Nuevo artículo 64.4, introducido en el Estatuto de los Trabajadores por la denominada Ley 'Riders'

Ley 12/2021, de 28 de septiembre, por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales (BOE de 29/09/2021).

El artículo único modifica el artículo 64 en lo relativo a los derechos de información y consulta de la representación legal de las personas trabajadoras, añadiendo un nuevo párrafo d) a su apartado 4, en el que se reconoce el derecho del comité de empresa a ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles.

2. Ley 15/2022 de igualdad de trato y no discriminación

Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación (BOE de 13 de julio de 2022).

También incluye expresamente el abordaje de la inteligencia artificial. Específicamente se aborda la igualdad de trato y no discriminación en el ámbito de la IA y los mecanismos de toma de decisión automatizados.

El Art. 3.1.o) establece expresamente que se aplicará en los siguientes ámbitos: *Inteligencia Artificial y gestión masiva de datos, así como otras esferas de análoga significación.*

El art. 23.3 sobre la IA y mecanismos de toma de decisión automatizados dispone que *Las administraciones públicas y las empresas promoverán el uso de una Inteligencia Artificial ética, confiable y respetuosa con los derechos fundamentales, siguiendo especialmente las recomendaciones de la Unión Europea en este sentido.*

3. V Acuerdo para el Empleo y la Negociación Colectiva

En el ámbito de la negociación colectiva, el **V Acuerdo para el Empleo y la Negociación Colectiva (AENC)** aborda expresamente la necesidad de regular usos y limitaciones de la IA a través de la

Resolución de 19 de mayo de 2023, de la Dirección General de Trabajo, por la que se registra y publica el V ANEC.

En concreto, establece en el Capítulo XVI, III lo siguiente en torno a *la Inteligencia Artificial (IA) y garantía del principio de control humano y derecho a la información sobre los algoritmos:*

La IA de manera progresiva tendrá un impacto significativo en el mundo laboral y, si no se hace un uso correcto y transparente, podría llevar a adoptar decisiones sesgadas o discriminatorias relativas a las relaciones laborales.

En línea con lo previsto en el Acuerdo Marco Europeo sobre digitalización, el despliegue de sistemas de IA en las empresas deberá seguir el principio de control humano respecto a la IA y ser seguro y transparente. Las empresas facilitarán a los representantes legales de las personas trabajadoras información transparente y entendible sobre los procesos que se basen en ella en los procedimientos de recursos humanos (contratación, evaluación, promoción y despido) y garantizarán que no existen prejuicios ni discriminaciones.

La negociación colectiva debe desempeñar un papel fundamental estableciendo criterios que garanticen un uso adecuado de la IA y sobre el desarrollo del deber de información periódica a la representación de los trabajadores.

El despliegue de sistemas de IA en las administraciones públicas debe seguir igualmente el principio de control humano y ser seguro y transparente. En base a lo anterior, las Confederaciones firmantes de este Acuerdo instamos al Gobierno para que facilite a los interlocutores sociales, a través de los órganos de participación institucional, la información suficiente que garantice la transparencia digital y algorítmica, en especial de aquellas fórmulas que configuran los aplicativos vinculados a las relaciones laborales y a la protección social.

4. Convenios colectivos que regulan expresamente usos y limitaciones de la IA

Aunque de manera lenta, ya hay algunos convenios colectivos que regulan los usos y limitaciones de la IA, como, por ejemplo, los siguientes:

4.1 XIX Convenio Colectivo de la Industria Química (BOE de 17/02/2025):

El 17 de febrero de 2025 se publicó en el BOE **la Resolución de 6 de febrero de 2025, de la Dirección General de Trabajo, por la que se registra y publica el XXI Convenio colectivo general de la industria química.**

En su art. 10 regula las nuevas tecnologías e inteligencia artificial. Entre otras cuestiones, establece lo siguiente:

> *1. Cuando en una empresa se introduzcan nuevas tecnologías que pueden suponer para las personas trabajadoras modificación sustancial de condiciones de trabajo, o bien un periodo de formación o adaptación técnica no inferior a un mes, se deberán comunicar las mismas con carácter previo a los representantes de las personas trabajadoras (RLT) en el plazo suficiente para poder analizar y prever sus consecuencias en relación con: empleo, salud*

laboral, formación y organización del trabajo, aspectos éstos sobre los que deberán ser consultados.

2. Asimismo, se facilitará a las personas trabajadoras afectadas la formación adecuada y precisa para el desarrollo de su nueva función.

En el supuesto de que las nuevas tecnologías que se vayan a introducir se basen en sistemas de inteligencia artificial (IA), la información que deberá facilitarse a la RLT aludirá, al menos, a las siguientes cuestiones sobre el Sistema concreto de IA a implantar.

Objetivos y razones para la implantación del sistema de IA.

Evaluación de los posibles impactos en el empleo, especificando los puestos de trabajo que puedan verse afectados.

Análisis de los posibles cambios en las condiciones de trabajo.

Parámetros, reglas e instrucciones en los que se basen los algoritmos o sistemas de IA que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles, al objeto de evaluar su impacto sobre el empleo y las condiciones de trabajo.

También se establece que: La introducción de nuevas tecnologías o sistemas de IA comportará, si procede, la actualización de la evaluación de riesgos laborales; Los sistemas de IA garantizarán que no se produzcan prejuicios ni discriminaciones y que el principio rector del mismo deberá ser el control humano.

Las modificaciones sustanciales debidas a la introducción de sistemas de IA se establecen expresamente en el convenio. De producirse modificación sustancial de condiciones de trabajo (MSCT) como consecuencia de la implantación de nuevas tecnologías y sistemas de IA, deberá seguirse el trámite del artículo 41 del ET y lo establecido en el artículo 28.5 del Convenio sobre MSCT.

4.2 XXV Convenio colectivo del sector de la banca:

Resolución de 20 de diciembre de 2024, de la Dirección General de Trabajo, por la que se registra y publica el XXV Convenio colectivo del sector de la banca (BOE de 1 de enero de 2025).

Regula en el art. 80.5 sobre Derechos digitales, la IA:

Las nuevas herramientas basadas en algoritmos pueden aportar valor hacia una gestión más eficiente de las Empresas, ofreciendo mejoras en sus sistemas de gestión. Sin embargo, el desarrollo creciente de la aportación de la tecnología requiere de una implantación cuidadosa cuando se aplica en el ámbito de las personas. Por ello, las personas trabajadoras tienen derecho a no ser objeto de decisiones basadas única y exclusivamente en variables automatizadas, salvo en aquellos supuestos previstos por la ley, así como derecho a la no discriminación en relación con las decisiones y procesos, cuando ambos estén basados

únicamente en algoritmos, pudiendo solicitar, en estos supuestos, el concurso e intervención de las personas designadas a tal efecto por la Empresa, en caso de discrepancia.

Las Empresas informarán a la RLPT sobre el uso de la analítica de datos o los sistemas de inteligencia artificial cuando los procesos de toma de decisiones en materia de recursos humanos y relaciones laborales se basen, exclusivamente en modelos digitales sin intervención humana. Dicha información, como mínimo, abarcará los datos que nutren los algoritmos, la lógica de funcionamiento y la evaluación de los resultados.

4.3 Convenio colectivo para los establecimientos financieros de crédito:

Resolución de 3 de julio de 2024, de la Dirección General de Trabajo, por la que se registra y publica el Convenio colectivo para los establecimientos financieros de crédito (BOE de 17 de julio de 2024).

Regula expresamente la IA en el art. 35 sobre los Derechos digitales ante la IA. En concreto, el apartado 5 estable:

Las nuevas herramientas basadas en algoritmos pueden aportar valor hacia una gestión más eficiente de las Empresas, ofreciendo mejoras en sus sistemas de gestión. Sin embargo, el desarrollo creciente de la aportación de la tecnología requiere de una implantación cuidadosa cuando se aplica en el ámbito de las personas. Por ello, las personas trabajadoras tienen derecho a no ser objeto de decisiones basadas única y exclusivamente en variables automatizadas, salvo en aquellos supuestos previstos por la Ley, así como derecho a la no discriminación en relación con las decisiones y procesos, cuando ambos estén basados únicamente en algoritmos, pudiendo solicitar, en estos supuestos, el concurso e intervención de las personas designadas a tal efecto por la Empresa, en caso de discrepancia.

Las Empresas informarán a la RLT sobre el uso de la analítica de datos o los sistemas de inteligencia artificial cuando los procesos de toma de decisiones en materia de recursos humanos y relaciones laborales se basen exclusivamente en modelos digitales sin intervención humana.

Dicha información, como mínimo, abarcará los datos que nutren los algoritmos, la lógica de funcionamiento y la evaluación de los resultados.

5. Futura Ley para el buen uso y gobernanza de la Inteligencia Artificial

Consejo de Ministros de 11 de marzo de 2024. Anteproyecto de Ley para el buen uso y la gobernanza de la Inteligencia Artificial.

El Anteproyecto de Ley, que se tramitará por la vía de urgencia, seguirá ahora los trámites preceptivos antes de volver al Consejo de Ministros para su aprobación definitiva como Proyecto de Ley y envío a las Cortes Generales para su aprobación.

Con el objetivo de que la Unión Europea disponga de un marco legal común para el desarrollo, comercialización y uso de sistemas de IA que eviten los riesgos para las personas, el reglamento

prohíbe determinados usos maliciosos de la IA, introduce obligaciones más rigurosas para sistemas considerados de alto riesgo y establece unos requisitos mínimos de transparencia para el resto.

Las prácticas prohibidas entraron en vigor el 2 de febrero de 2025 y desde el 2 de agosto de 2025 se podrán sancionar mediante multas u otras medidas adicionales (requerir su adaptación al sistema para que sea conforme, impedir que se comercialice…) aplicando el régimen sancionador que incorpora el Anteproyecto de Ley, dentro de las horquillas que fija el reglamento europeo. A modo de ejemplo, son prácticas prohibidas (entre otras muchas) Inferir emociones en centros de trabajo o educativos como método de evaluación para promoción o despido laboral, salvo por razones médicas o de seguridad.
A futuro será muy interesante ver en qué se materializa la futura Ley para el buen uso y gobernanza de la IA. El Consejo de Ministros aprobó el pasado 11 de marzo de 2024 el Anteproyecto de Ley para el buen uso y la gobernanza de la Inteligencia Artificial, aunque habrá que esperar a ver qué sucede durante la tramitación parlamentaria.

Referencias bibliográficas

Ley 12/2021, de 28 de septiembre, por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales. BOE nº 233, de 29/09/2021. Entrada en vigor: 30/09/2021.

Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación. BOE nº 167, de 13 de julio de 2022, páginas 98071 a 98109. https://www.boe.es/eli/es/l/2022/07/12/15

Resolución de 19 de mayo de 2023, de la Dirección General de Trabajo, por la que se registra y publica el V Acuerdo para el Empleo y la Negociación Colectiva. BOE nº 129, de 31 de mayo de 2023, páginas 75426 a 75447. Ministerio de Trabajo y Economía Social. https://www.boe.es/eli/es/res/2023/05/19/(2)

Resolución de 6 de febrero de 2025, de la Dirección General de Trabajo, por la que se registra y publica el XXI Convenio colectivo general de la industria química. BOE nº 41, de 17 de febrero de 2025, páginas 21869 a 22010. Ministerio de Trabajo y Economía Social. https://www.boe.es/eli/es/res/2025/02/06/(1)

Resolución de 20 de diciembre de 2024, de la Dirección General de Trabajo, por la que se registra y publica el XXV Convenio colectivo del sector de la banca. BOE nº 1, de 1 de enero de 2025, páginas 597 a 664. Ministerio de Trabajo y Economía Social. https://www.boe.es/eli/es/res/2024/12/20/(6)

Resolución de 3 de julio de 2024, de la Dirección General de Trabajo, por la que se registra y publica el Convenio colectivo para los establecimientos financieros de crédito. BOE nº 172, de 17 de julio de 2024, páginas 90000 a 90024. Ministerio de Trabajo y Economía Social. https://www.boe.es/eli/es/res/2024/07/03/(7)

Consejo de Ministros de 11 de marzo de 2025. Anteproyecto de Ley para el buen uso y la gobernanza de la Inteligencia Artificial
https://www.lamoncloa.gob.es/consejodeministros/referencias/Paginas/2025/20250311-referencia-rueda-de-prensa-ministros.aspx#artificial

La importancia de la Gobernanza del dato en la empresa

Joaquín García Onrubia, Dirección de Digitalización e IA en Enagas.
Colaborador del Proyecto IA+Igual.

Los datos son un activo estratégico fundamental para implantar la IA en la empresa y en el ámbito laboral. Como tal activo, es necesario garantizar su calidad y confiabilidad, establecer los roles y responsabilidades clave para su correcta gestión y facilitar un acceso seguro y ético para su desarrollo y uso. Las empresa se pregunta, en este contexto digital, cómo medir el impacto de la gobernanza de los datos en la eficiencia operacional y la innovación partiendo del cumplimiento normativo.

1. Activo estratégico fundamental

En el panorama actual, el dato ha trascendido su papel tradicional para convertirse en un activo crítico y un recurso estratégico fundamental para cualquier organización. De la misma manera que se gestiona el capital financiero, la correcta gestión de los datos es esencial para evitar la pérdida de oportunidades y el desperdicio de recursos. Contar con datos fiables y de alta calidad sobre aspectos esenciales como clientes, productos y operaciones es lo que, realmente, permite tomar decisiones informadas y sólidas que guían el rumbo de la compañía. Para obtener valor de estos activos, es indispensable planificar y ver el dato como un producto que se crea y se gestiona activamente. Las decisiones tecnológicas deben estar alineadas con los requisitos de datos que, a su vez, responden a la estrategia de negocio. Esta perspectiva es el cimiento sobre el cual se pueden construir iniciativas avanzadas, incluida la implementación exitosa de la inteligencia artificial en los distintos espacios del ámbito laboral.

La Gobernanza del dato emerge como el elemento clave para maximizar el valor de los datos y facilitar la toma de decisiones alineadas con los objetivos de negocio. Constituye el marco organizativo necesario para definir la estrategia, los objetivos y las políticas que aseguren una gestión eficaz de los datos corporativos. Y es, además, el fundamento corporativo que, de manera independiente a los sistemas específicos, ayuda a proteger la calidad de los datos y a soportar procesos de negocio eficientes.

La gobernanza proporciona los principios, políticas, procesos, estructuras, métricas y la supervisión necesaria para guiar todas las actividades de gestión de datos en la organización, garantizando que la alta dirección disponga de la información necesaria para tomar decisiones informadas y lograr los objetivos de la organización. El valor del dato reside, precisamente, en cómo se utiliza para cumplir dichos objetivos de negocio, lo que subraya la importancia de un gobierno del dato robusto y enfocado al valor. Este marco asegura, además, que el uso de los datos sea ético y adecuado, cumpliendo con toda la normativa aplicable.

Para que una organización pueda realmente capitalizar el potencial de sus datos y avanzar hacia una visión orientada a los datos (data-driven), es fundamental posicionar la Gobernanza del dato como una función transversal y estratégica. No es una imposición externa, sino que surge de las necesidades del propio negocio, adaptándose a la estructura y cultura existentes en la organización. Es un pilar indispensable en la transición de un estado donde

los datos están dispersos, son heterogéneos y de baja calidad, a otro donde se dispone de productos de datos estandarizados, accesibles, reutilizables y gobernados. Esta transformación requiere un liderazgo comprometido y la participación activa de todos los niveles de la organización. La gobernanza proporciona el marco necesario para que la gestión de datos se alinee de manera efectiva con las prioridades del negocio y los stakeholders. Su integración en los proyectos estratégicos de la compañía es crucial.

2. Garantizar la calidad y confiabilidad

La efectividad de cualquier iniciativa que utilice datos y, en particular, la precisión de los modelos de inteligencia artificial, dependen de manera crítica de la calidad y confiabilidad de la información con la que trabajan. Para que los sistemas de IA generen resultados precisos e insights valiosos, necesitamos una base de datos sólida y de alta calidad.

La Gobernanza del dato actúa como un motor esencial para impulsar las iniciativas de mejora de la calidad del dato. Esta es una disciplina fundamental de control que abarca la evaluación, gestión, uso, mejora, monitorización y protección de nuestra información organizacional. Un marco de gobernanza bien establecido define la estrategia y las políticas de gestión de datos, y es crucial para proteger la calidad del. Es el mecanismo mediante el cual se asigna la responsabilidad sobre la mejora de la calidad del dato, haciendo transparente quién debe actuar para resolver problemas, especialmente en datos maestros. Para que sea efectiva, la gobernanza debe adaptarse a nuestra estructura y cultura, fomentando una gestión del dato que considere los intereses de la empresa y de sus grupos de interés. Nuestros datos son activos críticos que requieren ser gobernados y protegidos, y la gobernanza proporciona la autoridad y el control necesarios para coordinar personas, procesos y tecnologías hacia este fin. Las iniciativas de gobierno del dato tienen que incluir explícitamente la calidad como uno de sus pilares fundamentales.

Una base de datos deficiente o de baja calidad tiene un impacto negativo significativo en nuestra capacidad para emprender nuevos procesos y proyectos, incluyendo las ambiciosas iniciativas de IA. Los datos inexactos, incompletos o desactualizados son inherentemente riesgosos y pueden llevar a decisiones equivocadas. Las brechas de información -esa diferencia entre lo que sabemos y lo que necesitamos saber- pueden tener un impacto profundo en nuestra eficiencia operativa y rentabilidad. La proliferación de datos no controlados y las relaciones complejas entre sistemas minan la eficiencia y, sobre todo, reducen la confianza en los datos disponibles. Para que nuestros análisis, especialmente los que se basan en modelos de IA, sean fiables, los datos que los alimentan deben ser fiables. En el contexto de los flujos de datos que nutren la IA, la baja calidad se convierte en una preocupación aún mayor. Los modelos avanzados, como los Grandes Modelos de Lenguaje (LLM por su sigla en inglés), aplican una enorme capacidad computacional a los datos, lo que significa que cualquier deficiencia en la calidad del dato se amplifica en el resultado final, generando insights erróneos o modelos poco fiables. Una gestión inadecuada de la calidad puede, de hecho, resultar en una pérdida de foco estratégico y una baja apreciación del valor del dato por parte del negocio.

La mejora de la calidad del dato es un proceso que abordamos de forma estructurada, basándonos en la definición de reglas claras y la ejecución de planes de remediación. Iniciamos definiendo qué constituye un dato de alta calidad y estableciendo una estrategia de calidad. Identificamos los datos más críticos y las reglas de negocio asociadas. Luego, evaluamos la calidad existente para detectar problemas, priorizarlos y analizar sus causas raíz. Consideramos el dato como un producto de nuestros procesos y aplicamos técnicas de mejora de calidad similares a las de la manufactura. Un paso crucial es la definición de reglas de calidad, tanto técnicas como funcionales. Estas reglas incluyen umbrales de aceptabilidad. Implementamos validaciones en las aplicaciones

y ejecutamos controles de calidad, monitorizando continuamente los resultados. Si algún indicador cae por debajo del umbral, activamos un plan de remediación, colaborando entre los responsables de negocio y los equipos técnicos. El objetivo es lograr mejoras medibles, como una mayor completitud de los datos, y demostrar su valor de negocio para justificar la inversión en calidad.

Todas estas acciones, impulsadas por la gobernanza, tienen como fin último aumentar la confianza y la transparencia sobre los datos que tenemos disponibles para su uso, especialmente en sistemas de IA. Buscamos que nuestros datos sean accesibles, relevantes, completos, precisos, consistentes, utilizables, significativos y comprendidos. La gobernanza mejora la confianza y transparencia no solo a través de la mejora de la calidad, sino también implementando mecanismos de seguridad. Al hacer los datos más fiables, consistentes y accesibles mediante un gobierno efectivo, construimos una base sólida para la adopción de IA, permitiendo entrenar sistemas con datos de mayor calidad. Esto, a su vez, conduce a insights y decisiones basadas en IA que son más precisas y, por lo tanto, más confiables.

3. Establecer roles y responsabilidades

Establecer un marco sólido para la gestión de nuestros datos es un paso fundamental, especialmente cuando pensamos en el uso de la inteligencia artificial. Para lograrlo de manera efectiva, es absolutamente crucial definir de forma clara los roles y responsabilidades en el ámbito del gobierno del dato.

La Gobernanza del dato aborda este problema directamente al formalizar la asignación de responsabilidades y la autoridad, lo que impulsa la mejora en la calidad de nuestra información, transitando de un estado con responsabilidades poco definidas a un marco de gobierno estructurado con roles y responsabilidades bien delineados.

La transparencia que se logra al definir estos roles y responsabilidades nos permite algo esencial: localizar rápidamente a las personas clave para resolver cualquier problema relacionado con los datos de una forma mucho más eficiente. Imaginemos a alguien nuevo en un equipo que necesita ayuda con un dato maestro importante para un proceso complejo. Si contamos con una estructura de gobierno del dato bien definida, la persona recién incorporada podrá identificar, sin lugar a duda, quién es el contacto adecuado que, gracias a su conocimiento especializado, puede analizar el problema con precisión y tomar las medidas necesarias para mejorar la calidad de esos datos. Herramientas como las matrices de responsabilidad -por ejemplo, el modelo RACI (Responsable, Corresponsable, Consultado, Informado)-, son muy útiles para detallar quién tiene cada tipo de implicación, lo que también ayuda a separar funciones y garantizar la seguridad de la información. Un principio clave aquí es la rendición de cuentas: la organización se responsabiliza de la información que controla y designa a individuos específicos para asegurar que se cumplan las políticas. La concienciación, la propiedad y la rendición de cuentas son los motores principales para involucrar a todo el personal en las iniciativas y procesos de gestión del dato.

Dentro de este marco de gobierno, hemos identificado y definido funciones y responsabilidades específicas para dos roles que consideramos pilares: el Data Owner (dueño de los datos) y el Process Data Ower (dueño de los datos del proceso).

El dueño de los datos es quien tiene la visión más amplia y estratégica sobre todos los procesos de un dominio de negocio. Es la persona que comprende mejor cómo funcionan los procesos dentro de su dominio y cómo se interrelacionan entre sí. Las responsabilidades principales son las siguientes:

- Garantizar que los datos de su ámbito estén correctamente definidos, gestionados y que se puedan relacionar adecuadamente con otros conjuntos de datos.

● Identificar las necesidades de negocio en cuanto a datos, tanto las actuales como las futuras, lo que lleva a la definición de productos de datos.

● Asegurar que el Glosario de negocio se mantenga actualizado para su dominio, participando activamente en su flujo de gestión.

● Determinar el nivel de seguridad que deben tener los datos bajo su control (estableciendo si son confidenciales, internos o públicos), lo cual es fundamental para controlar el acceso de los usuarios al catálogo de datos.

● Validar el acceso a los datos y establecer las condiciones de uso.

● Identifica los datos maestros y de referencia. Tiene la autoridad sobre el dominio completo que abarca, siendo quien ve y tiene autoridad sobre el "bosque" de datos.

● Es capaz de monitorizar en tiempo real quién accede a qué productos de datos y para gestionar directamente las listas de usuarios para dar de baja a quienes ya no necesitan ese acceso.

● En los flujos de trabajo, interviene en la identificación y definición de términos del glosario y, en relación con la calidad, identifica problemas, define reglas funcionales y colabora en el seguimiento y la elaboración de planes de mejora.

El PDO (siglas en inglés de Process Data Ower) es la persona con el conocimiento más detallado y profundo de un proceso específico y de los datos ligados a él. Su función principal es la gestión minuciosa de los datos dentro del contexto de ese proceso particular, lo que se compara con el cuidado individual de un "árbol".

Entre sus responsabilidades clave se encuentran:

● Garantizar que los datos de su proceso sean comprendidos por medio del Catálogo de Datos, colaborando en su enriquecimiento funcional y validación.

● Asegurar la calidad de los datos dentro de su proceso, lo que implica proponer y supervisar las acciones necesarias para mejorar esa calidad.

● También puede ser quien detecte y proponga nuevos productos de datos que surjan o estén relacionados con su proceso.

● Aunque la gestión formal del acceso a los datos de su proceso es responsabilidad del dueño de los datos, esta función puede ser delegada en el dueño de los datos del proceso.

● En el flujo del Catálogo, los PDO identifican las tablas relevantes dentro de los productos de datos, aportan descripciones funcionales y técnicas, y validan funcionalmente los datos catalogados. En el flujo de Calidad, son ellos quienes identifican los problemas específicos de su proceso, definen las reglas funcionales, hacen seguimiento de los resultados y participan activamente en la creación y ejecución de los planes de corrección.

● Al igual que el Data Owner, tienen la capacidad de monitorizar quién accede a los productos de datos asociados a su proceso y pueden gestionar las listas de usuarios para retirar el acceso a quienes no corresponda.

4. Facilitar un acceso seguro y ético

El establecimiento de un marco de Gobernanza del dato es esencial para las organizaciones y compañías. Este marco asegura que el acceso a los datos sea seguro y ético, lo cual es fundamental para la toma de decisiones y el logro de los objetivos estratégicos de la organización. Los datos son considerados activos críticos que requieren ser gobernados y protegidos.

La seguridad es un área principal del gobierno del dato, considerándose un proceso confiable como un activo organizacional. La seguridad no solo evita el acceso no autorizado, sino que también facilita el acceso adecuado. La definición clara de roles y responsabilidades dentro del marco de gobierno aporta transparencia, lo que

permite identificar rápidamente a las personas clave para resolver problemas relacionados con los datos de manera eficiente. Formalizar la asignación de responsabilidades y autoridad contribuye a mejorar la calidad de la información. Fomentar una cultura del dato entre los empleados es también un objetivo importante. Algunas áreas donde las organizaciones pueden tener menor madurez incluyen el desarrollo de esta cultura del dato y la capacidad de compartir y describir adecuadamente los activos de datos.

Es vital implementar mecanismos de seguridad, incluida la clasificación de los datos según su nivel de sensibilidad. Un rol de Data Owner establece esta clasificación, que se utiliza para asegurar que solo los usuarios apropiados puedan acceder y utilizar el catálogo de datos de acuerdo con sus roles.

Esta clasificación es crucial para la definición de los datos y ayuda a reducir el riesgo al limitar la proliferación de datos sensibles o confidenciales. Los resultados de la clasificación deben guardarse en un repositorio central para informar a los empleados sobre la sensibilidad de la información. Herramientas como el enmascaramiento o la encriptación son útiles para restringir el movimiento de datos sensibles. Es necesario definir y supervisar políticas para proteger los datos.

La existencia de procesos definidos para la validación y aprobación del acceso y las condiciones de uso de los datos. Dentro del marco de gobierno, el rol de Data Owner se encarga de validar el acceso y establecer las condiciones de uso de los datos. Para obtener acceso a los productos de datos, los usuarios deben enviar una solicitud a este rol, explicando el propósito de uso.

El Data Owner recibe un flujo de trabajo de aprobación y decide si comparte el producto de datos para el fin indicado, pudiendo imponer condiciones específicas, como un límite de tiempo o la aplicación de filtros. Si se rechaza la solicitud, se informa al usuario con los motivos. En caso de aprobación, el equipo técnico se encarga de gestionar la inclusión del usuario en el grupo pertinente.

La capacidad para monitorizar y gestionar quién accede a qué productos de datos. Además, pueden gestionar directamente las listas de usuarios para revocar accesos si es necesario. La monitorización de la autenticación y el comportamiento de acceso proporciona información sobre quién se conecta y accede a los activos de información, lo que ayuda a detectar transacciones inusuales o sospechosas que puedan requerir investigación.

El compromiso con el uso ético y adecuado de los datos en conformidad con la normativa aplicable. Las organizaciones se comprometen a garantizar que los datos se utilicen de manera ética y adecuada, cumpliendo con la legislación vigente. Se implementan mecanismos para asegurar que se conozcan, comprendan y cumplan los compromisos de la compañía, así como para coordinar las actividades que garantizan la seguridad y privacidad del dato. Las políticas de gobierno del dato deben estar alineadas con los requisitos legales, regulatorios y contractuales. Esto incluye principios como la responsabilidad, donde los controladores de datos deben poder demostrar el cumplimiento. Se destaca la necesidad de que las empresas proporcionen una seguridad razonable para los datos del consumidor. La rendición de cuentas es un motor clave para involucrar a todo el personal en las iniciativas de gestión del dato.

En el contexto regulatorio, la propuesta de la Ley de Gobernanza de Datos de la UE, Data Governance Act (DGA), plantea principios como la neutralidad para los proveedores de servicios de compartición de datos (DSS), prohibiendo el uso cruzado de datos y exigiendo separación legal. Los DSS deben proporcionar un alto nivel de seguridad para el almacenamiento y transmisión de datos no personales y adherirse a las normas de competencia. La DGA también regula las organizaciones de altruismo de datos, exigiendo que sean sin fines de lucro para ser

reconocidas, aunque la falta de requisitos claros para este reconocimiento se considera un problema. Es importante señalar que, a menudo, es difícil trazar límites claros entre datos personales y no personales, especialmente en "conjuntos de datos mixtos". La DGA también aborda el acceso a datos no personales por parte de autoridades policiales de terceros países. Se ha señalado que la inconsistencia en el vocabulario de la propuesta de la DGA podría dificultar su interpretación y cumplimiento. También se plantea una posible superposición de competencias entre distintos organismos reguladores europeos. La definición de términos como "dato", "titular del dato" y "usuario del dato" en la propuesta de la DGA plantea interrogantes fundamentales, y se ha realizado un análisis crítico de estas definiciones.

5. Eficiencia operacional e innovación

La gestión de datos, que incluye la gestión de metadatos y la gestión de la calidad bajo la supervisión del gobierno del dato, hace que los datos sean menos costosos de mantener y más fáciles de utilizar, lo que amplía las oportunidades para su explotación. La existencia de reglas, roles y responsabilidades claras en torno a los datos contribuye a que los procesos operativos funcionen de manera fluida, ahorrando tiempo. Las fases de preparación y análisis de datos pueden consumir una parte considerable del tiempo en los proyectos de analítica.

El gobierno del dato busca transformar un panorama de datos disperso y heterogéneo en un catálogo accesible de productos de datos de calidad y estandarizados, lo que acelera el acceso a la información y contribuye a eliminar los cuellos de botella en el análisis. Los procesos manuales en la gestión del dato pueden generar errores e inconsistencias, exacerbando los problemas; la documentación y catalogación manual, que requieren mucho tiempo, pueden representar una carga significativa. Un gobierno robusto, potencialmente asistido por tecnología, puede abordar estas ineficiencias manuales. Definir los requisitos de metadatos en función de la necesidad de comprender, acceder y compartir datos es un enfoque práctico que reduce el esfuerzo. La ingesta de datos con perfilado automático que proporciona metadatos a los analistas también mejora la eficiencia. Abordar los activos incompletos o desactualizados también aumenta la eficiencia analítica.

El objetivo fundamental es evolucionar desde un estado donde los datos están aislados, heterogéneos y son de baja calidad, con un gobierno centrado en sistemas, hacia un modelo con un amplio catálogo de productos de datos de calidad, estandarizados, fácilmente accesibles, altamente reutilizables y gobernados. La gobernanza de estos productos de datos, particularmente en capas estandarizadas, implica la definición de funcionalidades y técnicas, la aplicación de atributos contextuales, la corrección de errores, el monitoreo de la calidad, el establecimiento de controles de acceso y la clasificación de datos por sensibilidad.

Esto asegura que las tablas definidas sean altamente reutilizables por los usuarios, promoviendo su autonomía y garantizando la calidad y confiabilidad de los datos. La creación de un mapa de datos maestros de la compañía también facilita la transversalidad y la reutilización de los datos. Una gobernanza de datos robusta es un factor clave para la adopción exitosa de la IA.

La transición hacia un catálogo de datos gobernados busca facilitar un acceso rápido y eficiente a la información. Gobernar los productos de datos asegura su reutilización y promueve la autonomía de los usuarios al garantizar la calidad y confiabilidad de la información. La visión es que cada empleado pueda acceder a la información clave, que esté gobernada y sea fiable, y tenga la capacidad de generar y gestionar sus propios insights. Esto implica desarrollar la capacidad de los empleados para tratar, comprender y comunicar datos (Data Literacy), enfocándose en la autogestión, analítica, gobierno y compartición de datos. La

gobernanza refuerza la seguridad al tiempo que facilita un acceso adecuado.

La Gobernanza del dato se centra en identificar los datos más importantes (internos y externos) que están alineados con los Indicadores Clave de Rendimiento (KPIs) estratégicos y en mejorar la eficiencia en su adquisición, gobierno y uso. Se plantea colaborar con las áreas para identificar los KPIs clave alineados con el Plan Estratégico y apoyar su correcto gobierno, proporcionando las capacidades necesarias para un seguimiento eficiente y un análisis avanzado según se requiera. La visión del gobierno debe ser transversal y estratégica, identificando los datos principales de cada área alineados con KPIs estratégicos e implementando el gobierno para estos datos clave identificados. El gobierno del dato apoya explícitamente las iniciativas digitales y de IA al dirigir los esfuerzos hacia los datos más demandados y estratégicamente relevantes. Una gobernanza robusta de los datos es esencial para catalizar la adopción exitosa de la IA.

Una reflexión estratégica clave es la integración de la función de gobernanza en los proyectos estratégicos, aprovechando las nuevas herramientas disponibles. Las políticas de Gobernanza del dato se alinean con la estrategia y los objetivos del negocio, estableciendo un marco para la gestión del gobierno del dato, y busca posicionar este como algo transversal y estratégico, integrándolo en los proyectos clave. Las disciplinas fundamentales de gestión de datos, supervisadas por el gobierno del dato, como la gestión de riesgos, metadatos y calidad, son cruciales para la producción y el uso de datos, y deben ejecutarse con disciplina para dar soporte a las actividades analíticas.

6. Cultura orientada a los datos

Para que una organización logre aprovechar plenamente los datos como un activo corporativo, es esencial que su cultura aprenda a valorar tanto los datos mismos como las actividades relacionadas con su gestión. La Gobernanza del dato aporta valor al promover cambios en el comportamiento de las personas dentro de la organización. La adopción de prácticas como la arquitectura de datos depende en gran medida de la receptividad cultural hacia un enfoque arquitectónico. De manera similar, mejorar la calidad de los datos requiere modificar la forma en que las personas piensan y se comportan respecto a ellos. Al fomentar una cultura de gestión del dato, es crucial considerar no solo los intereses de la empresa, sino también las necesidades y expectativas de sus grupos de interés. Establecer una cultura para el manejo ético de los datos implica cumplir con la ley y asegurar el uso responsable de los datos en su análisis, interpretación y aprovechamiento; exige comprender las prácticas actuales, definir los comportamientos deseados, plasmarlos en políticas y un código de ética, y ofrecer capacitación y supervisión para garantizar su cumplimiento. Esta transformación cultural requiere un proceso formal de Gestión del Cambio Organizacional (OCM).

Un programa de Gobernanza del dato debe incluir actividades de gestión del cambio organizacional para educar a la organización e incentivar comportamientos que faciliten el uso estratégico de los datos. La necesidad de un cambio cultural es amplia, abarcando todas las responsabilidades del gobierno del dato, y se vuelve más crítica a medida que la organización madura sus prácticas de gestión. Para muchas organizaciones, este cambio cultural constituye un desafío significativo. Un principio fundamental de la gestión del cambio es que el cambio organizacional se basa en el cambio individual. Por muy precisa que sea la estrategia de gobierno del dato, ignorar la cultura reducirá las posibilidades de éxito. Por ello, el enfoque en la gestión del cambio debe ser una parte intrínseca de la estrategia de implementación.

Transformarse en una organización centrada en los datos significa que la organización valora los datos como un activo y los administra a lo largo de todo su ciclo de vida. El propósito de un programa de Alfabetización de Datos es proporcionar a los usuarios habilidades para que

puedan utilizar datos y maximizar su valor. Hay muchos ejemplos y antecedentes de planes de alfabetización de datos simples y escalables que tenían una baja penetración y adopción, subrayando la necesidad de reforzar este aspecto.

Potenciar la sensibilización acerca de los datos entre los empleados les permite procesar, comprender y comunicar este tipo de datos. Esta constituye una habilidad que se convierte en crucial para que puedan autogestionar, analizar, gobernar y compartir información. La capacitación en gestión de datos a todos los niveles de la organización está claro queincrementa la conciencia sobre las mejores prácticas y los procesos adecuados. Ser "lingüe en datos" implica la capacidad de articular el impacto de los datos en la estrategia y el éxito organizacional. Una evaluación de madurez de la gestión de datos (Data Management Maturity Assessment) puede ser útil para educar a los interesados acerca de conceptos, principios y prácticas de gestión de datos, aclarar roles y responsabilidades, destacar la necesidad de gestionar los datos como un activo crítico y además aumentar el reconocimiento de las actividades de gestión de datos en toda la organización.

La gestión de datos tiene un componente de innovación, que puede enfocarse en mejorar la toma de decisiones y la capacidad analítica a través de datos fiables y de mayor calidad. Un gobierno robusto de los datos, que asegura que los datos sean más fáciles de utilizar gracias a disciplinas como la gestión de riesgos, los metadatos y la calidad, es esencial para la adopción exitosa de la inteligencia artificial. Los profesionales que trabajan en analítica y ciencia de datos manejan información que describe diversos aspectos de personas y actividades. Explotar eficazmente estos datos para análisis avanzados y modelos de IA depende directamente de la calidad de los datos y de la habilidad de los profesionales para trabajar con ellos, capacidades que se fortalecen con una cultura de datos sólida y una alta alfabetización. Promover la conciencia sobre el rol y la importancia de los datos para la organización es el paso inicial en este cambio cultural, exigiendo que todos los empleados actúen de forma responsable con los datos con los que interactúan.

Día a día de la inteligencia artificial en la gestión de personas

Cristina Martín, directora de Talento y Personas de Optare Solutions.
Colaboradora del Proyecto IA+Igual.

La inteligencia artificial (IA) ha irrumpido en todos los ámbitos de nuestra vida y la gestión de personas no ha sido una excepción. Nos encontramos ante una nueva tecnología que, como ocurrió en anteriores revoluciones industriales, no podemos ignorar: adaptarse es imprescindible para seguir avanzando.

Quienes llevamos tiempo trabajando en el ámbito de Recursos Humanos hemos vivido una evolución tecnológica constante. Desde procesos manuales, incluso con máquina de escribir, hasta la progresiva incorporación de herramientas digitales que hoy son imprescindibles. Actualmente, es difícil encontrar un departamento de personas que no cuente con algún tipo de *software* para apoyar, total o parcialmente, sus responsabilidades. En este contexto, la Inteligencia Artificial se presenta como una tecnología más que debemos integrar, y cuyo potencial beneficio es innegable: bien implantada y gestionada, nos ayudará a hacer mejor nuestro trabajo, impactando positivamente en nuestros equipos y en los resultados de la organización.

IA en nuestro día a día

El objetivo fundamental de la gestión de personas sigue siendo el mismo: construir organizaciones con empleados satisfechos, motivados y comprometidos. Sabemos que este compromiso tiene una relación directa con los resultados empresariales. Desde esta premisa, todos los procesos del área pueden beneficiarse enormemente de la incorporación de la IA.

Uno de los impactos más inmediatos es la *automatización de tareas operativas*. Los procesos repetitivos pueden ser asumidos por sistemas inteligentes, permitiendo que los equipos de recursos humanos. centren su esfuerzo en labores de mayor valor estratégico. La inteligencia artificial destaca por su capacidad para procesar grandes volúmenes de información, estructurarla y analizarla, lo que nos permite tomar decisiones más fundamentadas y, probablemente, más acertadas que en escenarios donde la información es limitada o está dispersa. Este ha sido siempre uno de los grandes retos del área de talento: contar con información, información veraz y estructurada para poder realizar un buen análisis.

Si miramos las tendencias actuales en gestión de personas, encontramos un concepto clave: la *experiencia del empleado*. Desde el primer contacto en un proceso de selección hasta su desarrollo y trayectoria dentro de la organización, la experiencia debe ser cada vez más personalizada. En todo este ciclo de vida del empleado, la Inteligencia Artificial puede jugar un papel esencial, aportando información y soluciones adaptadas a las necesidades individuales.

Gestión ética y humana

Sin embargo, no podemos obviar los riesgos que supone una gestión basada exclusivamente en datos. Existe el peligro real de *deshumanizar las organizaciones* si olvidamos que detrás de cada dato hay personas: Juan, Sara, Gonzalo…

Si dejamos de considerar los matices, el contexto cultural y la historia individual de quienes forman parte de nuestros equipos, perderemos la esencia humana que define una organización saludable. La inteligencia artificial debe ser una herramienta al servicio de las personas, no un mecanismo que las sustituya en las decisiones más sensibles.

Otro aspecto crítico es *la vigilancia y la privacidad* en el entorno laboral. El uso de inteligencia artificial para monitorizar comportamientos debe abordarse con extremo cuidado, respetando la autonomía individual y evitando dinámicas de control excesivo y que pueda ser deshonesto.

En este nuevo contexto, la ética no puede ser una reflexión posterior: debe ser uno de los principios que guíen todo el diseño y la implantación de cualquier solución tecnológica en el ámbito de los recursos humanos.

Competencias del equipo de RR. HH.

Esta revolución tecnológica también transforma el perfil de quienes trabajamos en gestión de personas. Necesitaremos incorporar nuevas competencias: conocimientos tecnológicos, pensamiento crítico, habilidades analíticas, ética profesional e inteligencia emocional. Además, los equipos deberán ser más diversos y multidisciplinares que nunca para poder cuestionar, evaluar y mejorar continuamente los procesos en los que interviene la IA.

Será esencial establecer criterios desde el inicio que permitan detectar y corregir sesgos y prejuicios en los datos, evitando que los errores del pasado se amplifiquen usando la tecnología.

Gestionar este cambio representa un gran reto y una enorme responsabilidad: asegurarnos de que nuestras organizaciones sigan siendo, ante todo, humanas.

Glosario

A

Administración Algorítmica: Gestión de procesos administrativos de RRHH (nóminas, turnos, permisos, etc.) mediante sistemas de IA, orientados a la eficiencia operativa.

AI Act (Reglamento Europeo de IA): Marco legal aprobado en 2024 por la Unión Europea que clasifica los sistemas de IA según su nivel de riesgo. En RRHH, muchas aplicaciones se consideran de "riesgo alto" y están sujetas a requisitos de transparencia, trazabilidad y supervisión humana.

AI Ethics (Ética de la IA): Conjunto de principios y valores que orientan el diseño, desarrollo y uso responsable de la inteligencia artificial, especialmente en contextos sensibles como la gestión de personas.

AI Fairness 360 (AIF360): Biblioteca de IBM que ofrece métricas y técnicas de mitigación de sesgos algorítmicos antes y después del entrenamiento del modelo.

Ajuste Contextualizado de Modelos: Adaptación de los algoritmos a las realidades y culturas específicas de cada organización, evitando la aplicación ciega de estándares abstractos.

Ajuste de Umbrales por Grupo: Técnica de postprocesamiento que modifica los puntos de corte o criterios de clasificación por subgrupos, para compensar desigualdades.

Alfabetización Algorítmica: Conjunto de conocimientos y habilidades que permiten a los empleados comprender, evaluar y supervisar el funcionamiento de algoritmos en su entorno laboral.

Alfabetización Crítica en IA: Formación que no solo enseña el uso de herramientas, sino que capacita para comprender los sesgos, límites y derechos relacionados con la inteligencia artificial.

Algoritmo: Secuencia finita y ordenada de pasos o instrucciones definidos con precisión, que permite resolver un problema o ejecutar una tarea mediante un proceso automático.

Algoritmo de Matching: Sistema automatizado que compara las características de los candidatos con los requisitos de un puesto para determinar el grado de adecuación entre ambos.

Análisis Cualitativo: Enfoque metodológico basado en la interpretación de percepciones, experiencias y significados expresados por los actores implicados en el uso de IA en el trabajo. Se apoya en entrevistas, encuestas abiertas o focus groups.

Análisis de datos: Proceso de estudio de grandes volúmenes de datos con el objetivo de identificar patrones, tendencias o relaciones que faciliten la toma de decisiones fundamentadas.

Análisis de Sentimiento Organizacional: Evaluación automatizada del tono emocional de encuestas, mensajes o interacciones, para detectar señales de clima laboral o bienestar.

Análisis Predictivo: Método basado en el uso de datos históricos para predecir eventos futuros. En RR. HH. se usa, por ejemplo, para anticipar renuncias o diseñar planes de carrera más eficaces.

Anonimización de Datos: Técnica para eliminar la posibilidad de identificar a una persona a partir de los datos, protegiendo su privacidad durante el procesamiento o análisis.

Ansiedad Tecnológica: Reacción emocional adversa ante la incorporación de nuevas tecnologías, que puede mitigarse con metodologías inclusivas y empáticas de formación.

Aprendizaje Automático (Machine Learning): Subcampo de la IA que permite a los sistemas aprender de los datos y mejorar su rendimiento con el tiempo sin ser programados explícitamente. Es clave en aplicaciones predictivas como la identificación de talento o la detección de fuga.

Áreas Estratégicas de Impacto: Procesos clave donde la inteligencia artificial afecta directamente la equidad y diversidad: reclutamiento, desempeño, promoción, clima y representación.

Arquitectura Dinámica de Gobernanza: Modelo de gobernanza del dato que evoluciona continuamente para adaptarse a nuevos retos tecnológicos, normativos y sociales en el entorno laboral.

Arquitectura Ética de Datos: Diseño estructural que garantiza que los datos utilizados para IA cumplan principios éticos como la equidad, la protección de la privacidad y la inclusión.

Asistente Virtual: Herramienta digital, a menudo basada en IA, que interactúa con usuarios a través de lenguaje natural para brindar soporte o realizar tareas automatizadas. En RR. HH. se usa en on boarding, dudas frecuentes o formación.

Auditoría Algorítmica: Proceso sistemático de evaluación técnica, legal y ética de un sistema de IA para garantizar su equidad, transparencia, fiabilidad y cumplimiento con estándares normativos.

Auditoría Algorítmica Independiente: Evaluación externa realizada por entidades imparciales ajenas al equipo desarrollador o usuario del sistema, con el objetivo de verificar objetivamente su cumplimiento normativo y ético.

Auditoría Empírica: Evaluación basada en datos reales de uso que analiza el comportamiento efectivo de un sistema de inteligencia artificial, su rendimiento técnico, presencia de sesgos y nivel de explicabilidad.

Auditoría Ética de la IA: Revisión sistemática y documentada de los riesgos, efectos y decisiones de un sistema de IA desde una perspectiva ética y sociolaboral, considerando valores como la justicia, la autonomía y la inclusión.

Automatización: Sustitución de tareas humanas por sistemas tecnológicos que las ejecutan de forma autónoma. En Recursos Humanos, se traduce en procesos más rápidos y eficientes, como el envío de convocatorias o la gestión documental.

Autonomía Profesional Algorítmica: Capacidad de las personas para ejercer control y tomar decisiones informadas sobre su trayectoria laboral en entornos mediados por inteligencia artificial o algoritmos.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

B

Backtesting: Técnica de validación que evalúa la precisión de un modelo predictivo comparando sus resultados con datos históricos ya conocidos.

Brecha Digital: Desigualdad en el acceso, uso o aprovechamiento de las tecnologías digitales. En RR. HH., puede limitar el acceso equitativo a la formación o a procesos de selección digitalizados.

C

Caja Negra (Black Box): Término usado para describir sistemas de inteligencia artificial cuya lógica interna no es comprensible ni transparente para los usuarios, dificultando la rendición de cuentas.

Calibración de Scores por Grupo: Técnica de postprocesamiento que ajusta las puntuaciones generadas por un modelo según el grupo al que pertenece la persona candidata.

Canales Automatizados de Comunicación: Herramientas de mensajería, correo o intranet gestionadas mediante IA para distribuir información de forma segmentada y eficaz.

Capacidades Funcionales (IA inclusiva): Enfoque que describe lo que una persona puede hacer, en lugar de etiquetarla por limitaciones, para favorecer la inclusión en modelos algorítmicos.

Certificación Algorítmica: Proceso estructurado mediante el cual se acredita que un sistema de IA cumple con criterios éticos, técnicos, organizativos y regulatorios. En RR. HH., garantiza que los algoritmos respetan derechos laborales y principios de equidad.

Certificación Ética Escalable: Modalidad de certificación algorítmica que adapta progresivamente sus exigencias éticas y técnicas según el nivel de madurez digital de la organización, permitiendo una implementación por etapas.

Certificación Modular: Enfoque que permite certificar sistemas de inteligencia artificial por áreas funcionales específicas (por ejemplo, selección o formación), favoreciendo una adopción progresiva.

Certification Washing: Práctica de usar un sello ético como mero instrumento reputacional sin asumir un compromiso real con los valores que representa.

Ciclo de Vida del Dato: Secuencia completa de etapas que recorren los datos desde su captura hasta su almacenamiento y transformación, sobre las cuales debe aplicarse control técnico y supervisión.

Ciclo de Vida Ético del Algoritmo: Enfoque integral que asegura que todas las etapas del desarrollo y uso de la IA están guiadas por principios éticos desde el diseño hasta su retirada.

Ciclo Vicioso del Sesgo: Dinámica por la cual un sistema de inteligencia artificial reproduce sesgos presentes en los datos históricos y los refuerza en el tiempo, generando discriminaciones estructurales.

Clima Laboral Algorítmico: Evaluación del ambiente de trabajo mediante análisis de sentimiento, patrones de interacción y datos internos procesados por IA.

Código Ético del Dato: Declaración formal que establece principios y compromisos para el tratamiento justo, seguro y no discriminatorio de los datos laborales.

Comité de Revisión Algorítmica: Órgano interno, con participación plural, encargado de supervisar los sistemas de inteligencia artificial aplicados en Recursos Humanos para garantizar su equidad y transparencia.

Comité Transversal de Datos: Grupo multidisciplinar —con perfiles técnicos, legales, éticos y de ciberseguridad— que valida de forma colegiada los accesos y usos de datos en entornos automatizados.

Comités Éticos de IA: Órganos consultivos internos o externos encargados de evaluar la conformidad ética de los sistemas de IA antes, durante y después de su implementación, con capacidad para emitir recomendaciones y alertas.

Comités Éticos Interdisciplinares: Subtipo de comité ético de IA compuesto por profesionales de diferentes disciplinas (tecnología, derecho, ética, RR. HH., etc.), cuya diversidad garantiza una evaluación más completa y contextualizada del impacto algorítmico.

Comunicación Interna Personalizada: Aplicación de IA para adaptar contenidos, canales y formatos comunicativos a distintos perfiles dentro de la organización.

Comunicación Transformadora: Estrategia comunicativa que va más allá de lo informativo, permitiendo cuestionar la IA, proponer mejoras y construir confianza organizativa.

Concept Activation Vectors (CAVs): Técnica de interpretabilidad conceptual que permite verificar si un modelo responde a constructos relevantes del dominio profesional.

Confianza Comunicativa Algorítmica: Nivel de credibilidad y aceptación que los empleados otorgan a los mensajes generados o adaptados por IA dentro del entorno laboral.

Conjuntos de Validación por Colectivo: Subgrupos de datos diseñados para evaluar el rendimiento algorítmico por perfiles específicos, y así detectar y corregir sesgos asimétricos.

Consentimiento Informado, Explícito, Específico y Revocable: Autorización expresa y documentada mediante la cual una persona consiente el uso de sus datos en sistemas de IA, conociendo su finalidad, alcance y duración, y conservando el derecho a revocarlo en cualquier momento. Es un requisito clave en contextos laborales y automatizados.

Corresponsabilidad Algorítmica: Enfoque que distribuye la responsabilidad del uso de IA entre desarrolladores, proveedores, usuarios y supervisores humanos.

Criba Automatizada: Proceso mediante el cual un sistema de IA analiza grandes volúmenes de CVs para filtrar candidatos según criterios predefinidos, reduciendo la carga administrativa del área de selección.

Cribado Curricular Automatizado: Proceso en el que un sistema de IA filtra y evalúa CVs para preseleccionar candidatos, utilizando criterios definidos previamente.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Criterios de Equidad Algorítmica: Conjunto de parámetros que buscan asegurar que los algoritmos traten de forma justa a todos los grupos sociales, evitando impactos desproporcionados o discriminatorios.

Cultura del Dato: Mentalidad y práctica organizacional que prioriza el uso sistemático de datos como base para la toma de decisiones. En RR. HH., implica sustituir la intuición por evidencia empírica.

Cultura del Dato Ético: Evolución de la cultura del dato que incorpora principios de responsabilidad, inclusión, transparencia y respeto a los derechos, promoviendo un uso justo y legítimo de datos y algoritmos.

Chatbot Conversacional: Sistema de IA capaz de mantener interacciones con humanos mediante lenguaje natural. En RR. HH., se utilizan para responder preguntas frecuentes, gestionar procesos de on boarding o realizar entrevistas iniciales.

D

Dashboard de Impacto Algorítmico: Herramienta visual que permite monitorizar el comportamiento de sistemas de IA en términos de sesgos, resultados por subgrupos o cumplimiento ético.

Dashboards de Equidad: Paneles visuales que permiten observar de forma dinámica cómo varían las métricas de equidad entre distintos grupos o subgrupos.

Data Owner: Responsable funcional del dato, encargado de validar sus definiciones y asegurar su coherencia con los objetivos estratégicos del área de RR. HH.

Data Steward: Profesional encargado de la calidad, trazabilidad y limpieza de los datos durante todo su ciclo de vida.

Datos No Estructurados: Información que no se organiza en bases de datos tradicionales, como texto libre, imágenes, videos o audios. En RRHH, este tipo de datos se analiza con IA para extraer insights de encuestas, entrevistas o comunicaciones internas.

Datos Representativos: Conjunto de datos que reflejan adecuadamente la diversidad del entorno laboral. Son esenciales para evitar sesgos y garantizar equidad en los sistemas de inteligencia artificial.

Datos Sintéticos para la Inclusión: Datos generados artificialmente que simulan perfiles de colectivos vulnerables, utilizados para garantizar representación sin comprometer la privacidad.

Debiasing de Vectores: Técnica específica de NLP que elimina sesgos en representaciones semánticas de palabras o frases utilizadas en modelos de lenguaje.

Declaración de Conformidad Ética: Documento formal que certifica el cumplimiento de una organización con los requisitos éticos y técnicos definidos para un determinado nivel de madurez.

Deep Learning (Aprendizaje Profundo): Técnica avanzada de aprendizaje automático basada en

redes neuronales con múltiples capas, que permite reconocer patrones complejos. Es la base de sistemas actuales de reconocimiento de imágenes, voz y lenguaje en Recursos Humanos y la gestión de talento.

Derecho a la Explicación: Derecho de los individuos a recibir información comprensible sobre las decisiones tomadas por algoritmos que les afectan directamente.

Derecho a la Oposición Algorítmica: Posibilidad de impugnar una decisión automatizada y exigir una revisión humana cuando se considere injusta o errónea.

Derecho a la Rectificación: Derecho de los empleados a corregir errores que deriven de decisiones algorítmicas incorrectas o basadas en información incompleta o desactualizada.

Derecho a No Ser Evaluado Exclusivamente por una Máquina: Derecho recogido en el RGPD que protege a las personas frente a decisiones completamente automatizadas sin supervisión humana significativa.

Derechos Digitales en el Trabajo: Conjunto de derechos que protegen a los empleados en relación con el uso de tecnologías digitales en el entorno laboral (privacidad, información, supervisión humana).

Desagregación por Colectivos Vulnerables: Práctica de analizar el impacto y las métricas de desempeño de la IA separadamente para distintos grupos sociales con el fin de detectar y corregir desigualdades.

Desajuste Normativo-Algorítmico: Conflicto entre la lógica técnica de un sistema automatizado y los marcos legales o éticos que rigen el entorno laboral.

DigComp: Marco europeo de competencias digitales que agrupa habilidades en cinco áreas: alfabetización digital, comunicación, creación de contenido, seguridad y resolución de problemas. Se aplica en RR. HH. para diseñar programas de formación.

Dimensión Ética (de la Gobernanza): Pilar que exige que la IA respete principios de dignidad humana, equidad, inclusión y explicabilidad en su diseño y aplicación.

Dimensión Ética (en auditoría): Evaluación de aspectos como la autonomía, la equidad, la explicabilidad y la inclusión, considerando el impacto en derechos y valores humanos.

Dimensión Ética (IA en Recursos.): Marco que vela por principios como la equidad, la autonomía, la justicia y la explicabilidad en la aplicación de IA a personas.

Dimensión Jurídica (en auditoría): Análisis del cumplimiento con normativas como el RGPD, el AI Act y la legislación laboral nacional o europea.

Dimensión Legal (de la Gobernanza): Componente que asegura el cumplimiento de normativas como el GDPR y el AI Act, protegiendo los derechos fundamentales en el tratamiento de datos y uso de IA en RR. HH.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Dimensión Organizativa (de la Gobernanza): Hace referencia a la creación de estructuras internas (como comités éticos o de revisión algorítmica) que guían el uso responsable de la IA.

Dimensión Organizativa (en auditoría): Revisión del papel de los actores humanos, mecanismos de revisión y sistemas de gobernanza que rodean al uso de IA.

Dimensión Psicológica (de la Gobernanza): Perspectiva que analiza el impacto de la IA en el bienestar, la percepción de justicia y la motivación de los empleados.

Dimensión Técnica (de la Gobernanza): Abarca la validación, calidad, seguridad y trazabilidad de los datos que alimentan sistemas de IA, asegurando su funcionamiento legítimo y fiable.

Dimensión Técnica (en auditoría): Evaluación del comportamiento, rendimiento, calidad de datos, sesgos y robustez del sistema algorítmico.

Discriminación por Trayectoria Atípica: Riesgo de que la IA penalice carreras con interrupciones, informalidad o contextos no convencionales, perpetuando estigmas.

Diseño Ético por Defecto: Principio por el cual los sistemas de IA deben configurarse desde su origen para garantizar la equidad, la transparencia y la supervisión humana significativa.

Disparate Impact (DI): Indicador de discriminación estadística que mide si un subgrupo recibe un tratamiento significativamente diferente al grupo de referencia.

Diversidad en el Ciclo de Vida de la IA: Integración de principios de equidad e inclusión en todas las etapas de desarrollo, implementación y evaluación de sistemas de IA en RR. HH.

E

Efecto de Eliminación de Variables Sensibles: Supresión de atributos como género o discapacidad con la intención de evitar sesgos, que puede ocultar desigualdades reales al eliminar la capacidad de detectarlas.

Efecto de Paralización por Riesgo: Reacción organizativa ante la incertidumbre tecnológica que lleva a frenar el uso de IA por miedo a las consecuencias legales o reputacionales.

Efecto Piloto Perpetuo: Situación en la que un sistema de IA permanece en fase experimental sin consolidación, auditoría ni documentación estructurada.

Elementos de Inflexión Algorítmica: Factores que pueden cambiar el impacto de la IA de excluyente a inclusivo: transparencia, participación, supervisión humana y ajuste contextual.

Encuestas Abiertas: Instrumento de recogida de datos cualitativos que permite a los participantes expresar libremente sus opiniones sobre el uso de IA, sin opciones cerradas.

Entrevistas Semiestructuradas: Técnica cualitativa que combina preguntas abiertas con una guía temática flexible, permitiendo explorar en profundidad la experiencia individual de los usuarios de sistemas de IA.

Equalized Odds: Métrica de equidad que exige que la tasa de verdaderos positivos y falsos positivos sea igual entre distintos grupos sensibles.

Equidad Algorítmica: Principio que garantiza que las decisiones tomadas por sistemas de inteligencia artificial sean justas y no reproduzcan sesgos o desigualdades estructurales. En el ámbito de Recursos Humanos, es clave para asegurar procesos de selección, promoción y evaluación inclusivos.

Equidad Reforzada en IA: Estrategia que no solo evita sesgos, sino que implementa medidas activas para corregir desigualdades históricas mediante el uso justo de los algoritmos.

Estandarización Semántica: Proceso de alineación conceptual entre vocabularios técnicos distintos (p. ej., ESCO y catálogos nacionales), clave para la interoperabilidad.

Estrategia de Inserción Sociolaboral: Conjunto de políticas y acciones destinadas a facilitar el acceso y mantenimiento en el empleo de personas en situación de vulnerabilidad o exclusión social.

Estrategias de Mitigación del Sesgo: Conjunto de técnicas que permiten corregir desigualdades algorítmicas, actuando sobre los datos, el modelo o los resultados, sin comprometer la funcionalidad del sistema.

Evaluación de Impacto Algorítmico (AIA): Procedimiento para anticipar los efectos éticos, sociales y técnicos de un sistema de IA antes de su implementación, especialmente en contextos laborales.

Evaluación de Impacto en Protección de Datos (EIPD): Análisis obligatorio cuando un sistema automatizado afecta significativamente a las personas, que evalúa riesgos sobre la privacidad y define medidas de mitigación.

Evaluación de Impacto Ético-Laboral: Proceso sistemático que anticipa y valora las consecuencias éticas de implementar un sistema de inteligencia artificial en contextos laborales, analizando su impacto sobre la equidad, la autonomía o la dignidad profesional, con participación activa de los grupos afectados.

Explicabilidad Adaptada al Público: Principio según el cual las explicaciones ofrecidas por un sistema de IA deben ajustarse al perfil del receptor (trabajadores, técnicos, dirección, sindicatos, etc.), garantizando comprensión y utilidad.

Explicabilidad Algorítmica: Capacidad de un sistema de IA para ofrecer explicaciones comprensibles sobre cómo toma decisiones. Es fundamental para la transparencia, la rendición de cuentas y la supervisión humana en contextos como RR. HH.

Explicabilidad Diferenciada por Audiencia: Práctica de adaptar las explicaciones algorítmicas al perfil del receptor (comités sindicales, técnicos, usuarios finales), garantizando comprensión según el rol.

Exposición Injusta de Figuras Intermedias: Riesgo de responsabilizar a mandos medios de decisiones algorítmicas sin dotarlos de herramientas ni contexto para ejercer supervisión o apelación.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

F

Fairlearn: Herramienta que permite entrenar modelos con restricciones de equidad y visualizar resultados por subgrupos.

Fairness Constraints: Restricciones matemáticas que se incorporan al entrenamiento de modelos para penalizar comportamientos desiguales y reforzar criterios de equidad.

Fairness Indicators: Indicadores diseñados para evaluar y monitorizar el comportamiento equitativo de sistemas de IA, con especial atención a colectivos subrepresentados o vulnerables.

Fairness Individual: Enfoque que evalúa si personas con características similares reciben resultados similares, priorizando la equidad persona a persona.

Falsos Negativos (en equidad algorítmica): Casos en los que no se detecta un sesgo real, lo que permite que continúe una discriminación oculta.

Falsos Positivos (en equidad algorítmica): Casos en los que se detecta erróneamente un sesgo donde no lo hay, lo que puede inducir a intervenciones injustificadas.

Feedback Automatizado: Retroalimentación generada por sistemas de IA de forma inmediata, personalizada y continua, basada en datos de desempeño, comportamiento o participación. Se aplica, por ejemplo, en procesos de formación o evaluación laboral.

Focus Group: Metodología de discusión colectiva donde varios participantes comparten y debaten percepciones sobre el impacto de la IA en el entorno laboral.

Formación Diferenciada en IA: Diseño de programas formativos adaptados a distintos perfiles organizativos (técnicos, Recursos Humanos, directivos, plantilla) para garantizar un uso ético y comprensible de la IA.

Fragmentación del Trabajo: Efecto laboral de la automatización por el que las tareas se dividen en unidades mínimas y desconectadas, afectando la continuidad profesional.

G

Gamificación. Uso de elementos de juego en contextos no lúdicos, como la formación en RRHH, para aumentar la motivación, la retención de conocimiento y la implicación de las personas.

Gestión de Ciclo de Vida del Algoritmo: Supervisión de todas las fases de un sistema de IA, desde su diseño hasta su retirada, garantizando actualizaciones, revisiones y control de impactos.

Gobernanza del Dato: Conjunto de estructuras, roles, políticas y procesos destinados a garantizar la calidad, trazabilidad, representatividad y uso ético de los datos. En RRHH, protege la privacidad y asegura un tratamiento responsable de la información laboral.

Gobernanza Algorítmica: Marco organizativo que regula el diseño, desarrollo, implementación y supervisión de IA, asegurando su uso ético, transparente, justo y alineado con los derechos fundamentales.

Gobernanza del Dato y del Algoritmo: Modelo integral que coordina la supervisión tanto de los datos como de los sistemas algorítmicos, incorporando trazabilidad, control multidisciplinar y participación social.

Gobernanza Compartida de la IA: Modelo participativo en el que diversos actores organizativos —incluidos colectivos vulnerables— intervienen en el diseño, evaluación y control de los sistemas de inteligencia artificial.

Gobernanza Inclusiva: Enfoque de gobernanza del dato y de la IA que garantiza la representación activa de grupos diversos, promoviendo decisiones más justas y contextualizadas.

Gobernanza Transversal de la IA: Estrategia que implica a múltiples departamentos, niveles jerárquicos y perfiles profesionales en la supervisión ética del ciclo de vida de los sistemas de IA.

Gobernanza Pública-Privada de la IA: Colaboración estructurada entre administraciones públicas, empresas y sociedad civil para establecer marcos comunes de regulación, desarrollo y certificación de sistemas de IA.

Guías Internas de Compra de IA: Herramientas organizativas que orientan a responsables de RR. HH. sobre cómo evaluar, seleccionar y exigir transparencia a los proveedores de soluciones de IA.

H

Hipersegmentación Comunicativa: División excesiva de audiencias mediante IA, que puede llevar a la exclusión de grupos o pérdida de referencias comunes.

Hipersegmentación de Perfiles: Práctica algorítmica que divide a los candidatos en grupos muy específicos, reduciendo su movilidad o visibilidad en procesos de selección.

HRIS (Human Resources Information System): Sistema de información que permite almacenar y gestionar datos de RR. HH. Su evolución ha facilitado el paso de la gestión administrativa a un enfoque más analítico y estratégico.

I

IA Conversacional: Subtipo de IA diseñada para interactuar mediante lenguaje natural, usada en procesos de on boarding, atención al empleado o bots de bienvenida.

IA en Ciclo de Vida del Talento: Aplicación de la inteligencia artificial en todas las fases del recorrido profesional del empleado, desde la atracción hasta la desvinculación, con fines de eficiencia, personalización y análisis predictivo.

IA Explicable (Explainable AI): Conjunto de métodos y técnicas del campo de la inteligencia artificial diseñados para permitir que los sistemas generen explicaciones claras sobre su funcionamiento interno y decisiones.

IA Generativa: Rama de la inteligencia artificial que se centra en la creación de contenido nuevo (texto, imágenes, audio, o código) a partir de patrones aprendidos con grandes volúmenes de datos. En

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Recursos Humanos puede utilizarse, por ejemplo, para redactar descripciones de puesto o generar simulaciones de entrevistas.

IA Generativa Jurídica: Uso de modelos generativos para interpretar normativas o convenios, que requiere validación humana y jurídica para evitar errores críticos.

IA Laboral Inclusiva: Modelo de aplicación de inteligencia artificial en el trabajo que prioriza la equidad, representatividad, transparencia y participación en cada fase del ciclo de vida del talento.

IA para Todos (AI for All): Enfoque que promueve el desarrollo y uso de sistemas de IA accesibles, inclusivos y comprensibles para toda la población laboral, sin exclusiones.

Inclusión Activa en el Diseño de Datos: Estrategia que implica la participación directa de colectivos subrepresentados en la recolección y estructuración de datos para evitar exclusión desde el origen.

Inclusión Digital en Comunicación: Diseño de sistemas y contenidos accesibles para personas con baja familiaridad tecnológica o con barreras específicas de acceso.

Inclusión Sociolaboral Algorítmica: Uso de IA como herramienta para favorecer la empleabilidad y el acceso justo al trabajo de personas en situación de vulnerabilidad.

Indicador de Riesgo del Mánager: Métrica basada en la rotación histórica del equipo de un supervisor, utilizada como variable predictiva en modelos de retención.

Indicadores Clave de Desempeño (KPIs): Métricas utilizadas para evaluar el éxito de un área o proceso. En RR. HH, los KPIs han evolucionado desde indicadores descriptivos hasta métricas analíticas con apoyo de IA.

Indicadores de Equidad Algorítmica: Métricas utilizadas para evaluar el grado de equidad en los resultados de un sistema de IA, identificando sesgos o disparidades entre grupos.

Indicadores de Impacto Algorítmico: Métricas cuantitativas o cualitativas que permiten monitorear el efecto diferencial de la IA sobre distintos colectivos dentro de la organización.

Indicadores Éticos de Desempeño: Métricas diseñadas para medir el cumplimiento de principios como justicia, inclusividad o supervisión humana en los sistemas de IA.

Individual Conditional Expectation (ICE): Variante de PDP que representa el efecto de una variable sobre la predicción para cada individuo, permitiendo análisis más detallados.

Inferencia Emocional Algorítmica: Técnica mediante la cual la IA intenta deducir estados emocionales del usuario a partir de datos como texto, voz o biometría, sujeta a restricciones éticas y legales.

Infraestructura Crítica de Decisión: Conjunto de sistemas algorítmicos de inteligencia artificial que, al intervenir en decisiones sensibles como las laborales, requieren niveles elevados de transparencia y control.

Infraevaluación del Impacto Algorítmico: Falta de análisis técnico y social sobre las consecuencias reales de los sistemas automatizados implantados en procesos de RR. HH.

Infrarrepresentación Algorítmica: Ausencia o baja representación de colectivos vulnerables en los conjuntos de datos que entrenan sistemas de IA, lo que puede llevar a decisiones automatizadas injustas o excluyentes.

Inteligencia Artificial (IA): Disciplina tecnológica que desarrolla sistemas capaces de realizar tareas que tradicionalmente requerían inteligencia humana, como el aprendizaje, el razonamiento, la toma de decisiones o la comprensión del lenguaje.

Inteligibilidad Algorítmica: Grado en que un sistema de IA puede ser entendido en su funcionamiento y efectos por los distintos usuarios involucrados.

Interfaz Conversacional: Medio a través del cual los usuarios interactúan con sistemas de IA mediante lenguaje natural (escrito o hablado). En RR. HH., se usan en chatbots, asistentes virtuales y entrevistas automatizadas.

Interpretabilidad Algorítmica: Capacidad de entender cómo y por qué un modelo de IA toma decisiones, especialmente en sistemas opacos como los de tipo black-box.

Interseccionalidad Algorítmica: Consideración simultánea de múltiples ejes de discriminación (como género, edad, origen o discapacidad) al evaluar el impacto de la IA.

Interseccionalidad de los Sesgos: Acumulación de varios factores de exclusión (género, edad, discapacidad, origen, etc.) que se refuerzan mutuamente en los sistemas de IA.

In-Training (Entrenamiento Ético): Intervención durante el entrenamiento del modelo mediante penalizaciones, restricciones de equidad u optimización multiobjetivo para controlar sesgos mientras se ajustan los parámetros.

Inventario Dinámico de Algoritmos: Repositorio actualizado que recoge todos los sistemas de IA en uso, sus características, objetivos, responsables y estado de validación.

Itinerario de Mejora Ética Continua: Ruta progresiva de revisión, adaptación y mejora de los sistemas de IA, impulsada por auditorías, métricas y participación organizativa.

J

Justicia Algorítmica: Principio ético que exige que las decisiones tomadas por sistemas de IA sean imparciales, equitativas y auditables, evitando discriminaciones sistemáticas o sesgos hacia cualquier grupo.

L

Legitimidad Percibida: Grado en que los usuarios consideran justas y comprensibles las decisiones tomadas por sistemas de IA, independientemente de su precisión técnica.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Licencia Ética de Algoritmos: Modelo que condiciona el uso de ciertos sistemas de IA a la obtención de una certificación que garantice su conformidad con principios éticos mínimos.

LIME (Local Interpretable Model-agnostic Explanations): Técnica que genera modelos simples locales para explicar decisiones puntuales de modelos complejos, como perfiles de candidatos o puntuaciones de riesgo.

M

Marco Europeo de Certificación de IA en RR. HH.: Propuesta de estándar ético y técnico que armoniza criterios para la certificación de sistemas de IA en el ámbito de la gestión de personas en Europa.

Matching Algorítmico: Técnica que compara automáticamente el perfil de los candidatos con las características de un puesto para identificar coincidencias y sugerir compatibilidades.

Matriz de Riesgos Algorítmicos: Herramienta para identificar, clasificar y priorizar posibles impactos negativos de un sistema de IA, según su probabilidad y severidad.

Maximum Subgroup Loss: Métrica que identifica el peor rendimiento del modelo entre todos los subgrupos posibles, permitiendo diagnosticar riesgos ocultos.

Meta-Algoritmo (IA de segunda generación): Sistema automatizado que evalúa, regula o corrige las decisiones de otros algoritmos, requiriendo nuevos estándares de explicabilidad y control.

Microlearning: Estrategia formativa basada en contenidos breves y focalizados, que permite el aprendizaje flexible y personalizado. Es especialmente eficaz cuando se combina con tecnologías de IA.

Minimización de Datos: Principio por el cual solo deben recogerse los datos personales estrictamente necesarios para una finalidad específica, con el fin de reducir riesgos de uso indebido y proteger la privacidad.

Mitigación de Sesgos: Técnicas y estrategias aplicadas al diseño, entrenamiento y evaluación de algoritmos para evitar o corregir desigualdades en sus resultados.

Model Cards: Documentos estandarizados que describen objetivos, limitaciones, decisiones técnicas y riesgos de un modelo, promoviendo la transparencia.

Modelado de Abandono: Análisis predictivo del riesgo de rotación de personal, basado en variables como comportamiento, desempeño o entorno organizativo.

Modelo de Madurez Ética: Estructura escalonada de niveles (Bronce, Plata, Oro, Platino) que permite evaluar progresivamente el grado de alineamiento ético y técnico de una organización en el uso de IA.

Modelos Predictivos: Algoritmos que permiten anticipar eventos futuros a partir del análisis de datos históricos. En RR. HH. se usan para prever rotación, identificar talento o personalizar formaciones.

Modelos Prescriptivos: Algoritmos que no solo predicen lo que ocurrirá, sino que recomiendan acciones concretas para optimizar decisiones. En RR .HH., por ejemplo, proponen itinerarios de desarrollo individual.

Monitoreo Continuo del Desempeño Algorítmico: Seguimiento regular del funcionamiento real del sistema de IA en el entorno laboral para detectar desviaciones, sesgos o efectos no previstos.

Multiplicador de Sesgos u Oportunidades: Capacidad de la IA para amplificar tanto desigualdades como mejoras, dependiendo de su diseño, uso y supervisión.

N

Narrativa Algorítmica: Conjunto de mensajes internos generados o adaptados por sistemas de IA, que pueden influir en la cultura y percepción de la organización.

Narrativa Responsable sobre la IA: Comunicación clara, adaptada y ética sobre los usos, impactos y limitaciones de la IA dentro de la organización, especialmente dirigida a colectivos no técnicos.

Negociación Algorítmica: Uso de sistemas de IA en procesos de relaciones laborales como simulación de escenarios, interpretación de convenios o análisis de tensiones sindicales.

NLP (Procesamiento del Lenguaje Natural): Tecnología que permite a los ordenadores interpretar, entender y generar lenguaje humano. Se aplica en RR. HH. en sistemas como chatbots, análisis de encuestas abiertas o detección de emociones en texto.

No Discriminación Automatizada: Garantía de que los algoritmos no produzcan decisiones injustas contra personas por motivos de género, edad, discapacidad, etnia u otras condiciones personales.

O

Optimización Multiobjetivo: Técnica de ajuste de modelos que equilibra simultáneamente varios objetivos, como precisión y equidad, durante el entrenamiento del sistema.

Outliers Algorítmicos: Casos extremos o atípicos en los que el sistema presenta resultados inesperados o injustos, que podrían quedar ocultos en promedios agregados.

P

Pacto Ético sobre IA en el Trabajo: Compromiso formal asumido por una organización para aplicar la IA de forma transparente, justa, inclusiva y con participación de las personas afectadas.

Palancas de Inclusión Algorítmica: Factores clave que permiten transformar la inteligencia artificial en una herramienta de inclusión: representación crítica, sensibilización, formación y gestión de impacto.

Partial Dependence Plots (PDP): Técnica de interpretabilidad que muestra cómo cambia la predicción de un modelo al modificar el valor de una variable concreta.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Participación Activa de Colectivos Afectados: Inclusión deliberada de personas potencialmente impactadas por la IA en el diseño, despliegue y evaluación de los sistemas, como garantía de legitimidad y ajuste contextual.

Participación Deliberativa: Mecanismo estructurado que involucra activamente a trabajadores, sindicatos y otros actores en el diseño, supervisión y mejora de los sistemas de IA que afectan sus trayectorias profesionales, promoviendo la corresponsabilidad en su gobernanza.

Participación Plural en la Certificación: Involucramiento de colectivos diversos (incluyendo trabajadores, sindicatos o expertos externos) en la validación del uso ético de la IA.

People Analytics: Aplicación del análisis de datos a la gestión de personas. Incluye el uso de técnicas estadísticas, modelos descriptivos y herramientas de visualización para apoyar decisiones estratégicas en RR. HH.

People Analytics Predictivo: Subcampo del People Analytics que emplea análisis avanzado y modelos predictivos para anticipar comportamientos o necesidades futuras de los empleados, como la rotación, el compromiso o el desarrollo potencial.

Perfilado Algorítmico: Técnica mediante la cual los sistemas de IA generan perfiles individuales a partir de datos recopilados, con el objetivo de predecir comportamientos, adecuaciones o trayectorias. En RRHH se emplea en procesos de selección, formación o promoción.

Perfilado Preventivo: Modalidad específica de perfilado algorítmico orientada a clasificar a los trabajadores según su riesgo proyectado en ámbitos como la salud, la seguridad o la productividad, lo cual puede influir en asignaciones o restricciones laborales.

Perspectiva Multidisciplinar: Enfoque que integra dimensiones técnicas, legales, éticas y organizativas para analizar e implantar la IA en RR. HH de forma legítima y sostenible.

Política de Datos: Conjunto de directrices organizativas que regulan la recolección, tratamiento, almacenamiento y uso de datos laborales, con criterios de legalidad, equidad y transparencia.

Porcentaje de Falsos Positivos/Negativos por Grupo: Indicador específico que muestra si los sistemas de IA cometen más errores con determinados colectivos, revelando posibles sesgos estructurales.

Portal Algorítmico del Empleado: Plataforma interna que usa IA para personalizar contenidos, accesos y rutas de navegación, según el perfil y comportamiento del usuario.

Portales de Explicabilidad: Plataformas digitales que permiten a los empleados conocer de forma clara y comprensible cómo y por qué se toman decisiones automatizadas que les afectan.

Postprocesamiento (IA): Corrección aplicada después del entrenamiento del modelo, ajustando sus salidas o umbrales por grupo para garantizar resultados más equitativos.

Predictive Parity: Criterio de equidad que exige que la precisión de las predicciones positivas sea equivalente entre diferentes subgrupos.

Preprocesamiento (IA): Fase en la que se modifican los datos antes del entrenamiento del modelo, mediante técnicas como reponderación de clases o generación de perfiles sintéticos para reducir sesgos.

Prevención Predictiva con IA: Aplicación de IA para anticipar riesgos laborales, identificar patrones de accidentabilidad y generar recomendaciones de seguridad personalizadas.

Principio de Limitación de la Finalidad: Norma del RGPD (art. 5.1.b) que establece que los datos personales solo pueden ser utilizados para los fines expresamente consentidos, salvo justificación y documentación adicional.

Privacidad desde el Diseño: Enfoque proactivo que incorpora medidas de protección de datos personales desde la fase de concepción de los sistemas de IA, y no como un añadido posterior.

Procesadores Lógicos (LP): Enfoque que combina atributos sensibles (como género, edad u origen) para construir subgrupos interseccionales y evaluar el impacto algorítmico con mayor precisión.

Procesadores Multietapa (MP): Estrategia que combina técnicas de mitigación aplicadas en distintas fases del ciclo de vida del modelo (antes, durante y después del entrenamiento) para corregir sesgos desde múltiples ángulos.

Profecía Autocumplida Algorítmica: Riesgo de que una predicción generada por IA influya en la conducta organizativa de forma que provoque exactamente el resultado anticipado.

Proporción de Percepción de Justicia Algorítmica: Indicador subjetivo recogido mediante encuestas que mide el grado en que los trabajadores consideran justas las decisiones automatizadas que les afectan.

Proporcionalidad en el Uso de IA: Principio según el cual el uso de sistemas algorítmicos debe estar justificado, ser necesario y no excesivo en relación con los fines que se persiguen.

Prueba de Estrés Algorítmico: Simulación de condiciones extremas o atípicas de entrada para evaluar la robustez y estabilidad de un sistema de IA en contextos críticos.

R

Random Forest: Algoritmo de aprendizaje automático basado en múltiples árboles de decisión, utilizado en sistemas de scoring automatizado.

Regulación Proactiva: Enfoque normativo que busca anticipar riesgos y establecer marcos éticos y legales antes de que los problemas derivados del uso de la IA se materialicen.

Rendición de Cuentas Algorítmica: Principio según el cual debe haber responsables claros identificables por las decisiones y consecuencias derivadas del uso de sistemas automatizados.

Representación Crítica en Datos: Revisión activa de los datos utilizados en los sistemas de inteligencia artificial para asegurar la inclusión de colectivos infrarrepresentados y evitar exclusiones estructurales.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Reskilling: Proceso de adquisición de nuevas competencias por parte de una persona para adaptarse a funciones laborales distintas a las que venía desempeñando. Es clave ante cambios tecnológicos.

Responsabilidad Proactiva: Enfoque preventivo que obliga a anticipar y mitigar riesgos éticos o sociales antes de desplegar un sistema de IA.

Retroalimentación Descontextualizada: Información devuelta por un sistema de IA que no considera el entorno, la trayectoria o las condiciones del receptor, generando posibles malentendidos o juicios injustos.

Reutilización No Justificada de Modelos: Práctica de aplicar sistemas de IA a fines distintos a los previstos originalmente sin evaluación de compatibilidad ni nuevo consentimiento, contraviniendo el RGPD.

Reversibilidad Algorítmica: Capacidad de un sistema para que sus decisiones puedan ser revisadas, corregidas o revertidas por intervención humana.

Riesgo de Certificación Superficial: Peligro de que los procesos de certificación se limiten a cumplimientos formales sin impacto real sobre la equidad o la transparencia del sistema.

Riesgo de Exclusión Algorítmica: Posibilidad de que sistemas de IA ignoren o penalicen casos no normativos (como personas con discapacidad, víctimas de violencia o trayectorias atípicas).

Riesgo de Perfilado Ideológico: Potencial uso de datos para inferir posiciones políticas o sindicales, generando discriminación o vulneración de derechos colectivos.

Riesgo Residual Algorítmico: Nivel de incertidumbre o daño potencial que permanece tras haber aplicado medidas de mitigación sobre un sistema de IA.

S

Sandbox Ético: Entorno controlado de experimentación que permite evaluar el impacto de sistemas de IA antes de su implementación real, sin comprometer datos sensibles, y anticipar posibles efectos no deseados.

Sello Ético de Calidad: Reconocimiento formal que distingue a las organizaciones que implementan sistemas de IA responsables, transparentes y alineados con principios éticos en la gestión de personas.

Sensor Organizativo (IA): Uso de herramientas de IA para detectar de forma temprana señales sobre el estado emocional, clima o dinámicas internas dentro de una organización.

Sesgo Algorítmico: Distorsión sistemática e injustificada en los resultados de un sistema de IA, provocada por datos sesgados, modelos mal diseñados o falta de supervisión, que genera impactos desiguales sobre ciertos grupos, especialmente en contextos laborales.

Sesgo de Confirmación (en auditoría): Tendencia a interpretar los resultados algorítmicos en función de creencias previas, reforzando prejuicios institucionales.

Sesgo de Datos: Distorsiones en los resultados de los sistemas de IA provocadas por prejuicios o desigualdades presentes en los datos de entrenamiento.

Sesgo de Despliegue: Nuevas formas de sesgo que emergen cuando el sistema es utilizado en entornos reales, debido a bucles de retroalimentación o interacciones humanas.

Sesgo de Evaluación: Problema que ocurre al aplicar métricas de rendimiento sin desagregarlas por subgrupos, ocultando desigualdades que afectan a colectivos concretos.

Sesgo de Exclusión: Distorsión provocada por la omisión de variables relevantes o el uso de sustitutos que enmascaran factores sociales importantes.

Sesgo de Medición: Error que se produce cuando la variable objetivo utilizada por el modelo no refleja adecuadamente el fenómeno a predecir, generando resultados equívocos.

Sesgo de Representación: Tipo de sesgo que surge cuando ciertos colectivos están infrarrepresentados en los datos de entrenamiento, lo que reduce la precisión del sistema para esos perfiles.

Sesgo de Selección: Reproducción de patrones discriminatorios previos en los modelos de IA, como decisiones históricas de contratación sesgadas.

Sesgo del Modelo: Injusticias derivadas del diseño, parametrización o estructura técnica del algoritmo, que pueden favorecer o perjudicar a ciertos colectivos.

Sesgo Inconsciente: Juicio automático basado en estereotipos o experiencias previas que influye en decisiones sin que la persona sea plenamente consciente. Afecta procesos como selección, evaluación o promoción.

SHAP (SHapley Additive exPlanations): Técnica de interpretabilidad algorítmica basada en teoría de juegos que permite identificar la contribución de cada variable a una predicción específica.

Simulaciones Interactivas con IA: Herramientas formativas que permiten experimentar situaciones reales de toma de decisiones algorítmicas, favoreciendo la comprensión crítica.

Sistema ATS (Applicant Tracking System): Software utilizado para gestionar procesos de selección automatizada, incluyendo cribado y puntuación de currículums.

Sistema de Gestión Ética del Ciclo de Vida: Conjunto de políticas, herramientas y procesos que aseguran el cumplimiento ético durante todas las fases del sistema de IA, desde su diseño hasta su retirada.

Sistema de Recomendación Bidireccional: Arquitectura de IA que cruza información de vacantes y candidatos para generar recomendaciones mutuas, sin decisiones automatizadas.

Sistema de Revisión y Apelación: Mecanismo que permite a los empleados solicitar explicaciones, impugnar decisiones automatizadas y plantear mejoras en el uso de IA.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Sistema de Seguimiento de Candidatos (ATS): Software que automatiza la recepción, criba y evaluación inicial de candidaturas. Es una de las primeras aplicaciones de IA en RRHH y base de procesos de selección más ágiles y masivos.

Sistemas de Alto Riesgo (AI Act): Categoría regulatoria del AI Act que incluye aplicaciones de IA con gran impacto en los derechos de las personas. En RR. HH., comprende reclutamiento, evaluación de desempeño y promoción automatizada.

Sistemas de Tutoría Virtual Adaptativa: Plataformas que ofrecen recomendaciones y seguimiento de formación personalizados, ajustándose dinámicamente al progreso del usuario.

Sobre individualización del Riesgo: Enfoque que atribuye los riesgos laborales a características personales del trabajador, ignorando causas estructurales u organizativas.

Sostenibilidad tecnológica: Capacidad de una tecnología para perdurar en el tiempo de forma eficiente, minimizando impactos negativos y contribuyendo al bienestar social, económico y ambiental.

Statistical Parity Difference (SPD): Métrica que evalúa la diferencia entre la proporción de personas seleccionadas en cada grupo, en relación con una expectativa de paridad estadística.

Subgroup Fairness: Métrica de equidad que evalúa el comportamiento del modelo en subgrupos compuestos (por ejemplo, mujeres mayores racializadas) para detectar desigualdades interseccionales.

Supervisión Humana Significativa (SHS): Principio según el cual toda decisión automatizada relevante debe ser comprensible, cuestionable y corregible por una persona capacitada, con autoridad suficiente para intervenir y garantizar el respeto a los derechos de las personas afectadas.

T

Techos de Cristal Algorítmicos: Barreras invisibles que pueden reproducirse o amplificarse en sistemas de IA cuando penalizan trayectorias no lineales o perfiles subrepresentados.

Transformación Digital: Proceso de incorporación de tecnologías digitales en todas las áreas de una organización para mejorar su eficiencia, análisis de datos y capacidad operativa. En Recursos Humanos implica una transición hacia procesos más automatizados, predictivos y basados en evidencia.

Transformación Democrática Digital: Enfoque normativo de la transformación digital que exige que el uso de tecnologías esté guiado por principios de justicia social, equidad, transparencia y participación ciudadana, garantizando que el cambio tecnológico beneficie a todos los colectivos.

Transparencia Algorítmica: Capacidad de un sistema de IA para mostrar de forma clara y comprensible cómo y por qué toma decisiones automatizadas, permitiendo su interpretación por parte de personas usuarias y afectadas.

Transparencia Proactiva: Práctica de comunicar de forma accesible, anticipada y comprensible el funcionamiento y los impactos de los sistemas de IA que afectan a los empleados.

Trayectorias No Normativas: Recorridos profesionales que se alejan de los modelos lineales tradicionales y que pueden ser penalizados por sistemas automatizados mal entrenados.

Trazabilidad Algorítmica: Capacidad de registrar, reconstruir y documentar todas las etapas del funcionamiento de un sistema de IA, desde la entrada de datos hasta el resultado final, con el fin de permitir auditorías, supervisión y rendición de cuentas.

Trazabilidad de Datos: Capacidad de seguir el recorrido, origen y transformación de los datos a lo largo del tiempo. En RR. HH., garantiza el uso ético, controlado y transparente de la información personal.

Trazabilidad Documental: Registro sistemático de cada ajuste técnico y decisión normativa aplicada a un sistema de IA, para facilitar la auditoría y rendición de cuentas.

U

Upskilling: Proceso de actualización o mejora de competencias existentes para adaptarse a nuevas demandas dentro de un mismo rol profesional.

Upward Feedback Automatizado: Mecanismo por el cual los empleados pueden evaluar a sus superiores mediante sistemas de IA que analizan tendencias, comentarios y comportamientos, aportando objetividad a la gestión del liderazgo.

V

Validación Cruzada: Técnica estadística que evalúa la capacidad de generalización de un modelo aplicándolo a múltiples subconjuntos de datos, como k-fold o leave-one-out.

Validación Empírica: Proceso basado en evidencias observables que permite evaluar el comportamiento real de un sistema de IA en entornos laborales, verificando su impacto técnico y organizativo.

Validación Interseccional: Evaluación del rendimiento de un sistema de inteligencia artificial desagregando por múltiples categorías (género, edad, origen, etc.) para detectar sesgos combinados y asegurar equidad.

Valor Social de la Certificación: Impacto positivo que tiene el reconocimiento ético sobre la confianza pública, la reputación organizativa y la garantía de derechos laborales.

Vigilancia Algorítmica: Uso de tecnologías basadas en inteligencia artificial para monitorizar de forma constante la actividad, comportamiento o desempeño de personas. En Recursos Humanos y la gestión del talento, plantea riesgos para la privacidad y la autonomía laboral.

Vigilancia Emocional: Práctica controvertida en la que la inteligencia artficiail intenta inferir emociones del personal y trabajadores sin consentimiento claro, utilizando datos de comportamiento o biométricos.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

W

Warden AI: Plataforma centrada en auditoría continua de modelos, con alertas en tiempo real ante comportamientos sesgados.

What-If Tool (WIT): Herramienta visual e interactiva que permite explorar modelos de IA mediante análisis de sensibilidad y pruebas contrafactuales sin necesidad de código.

BIBLIOGRAFÍA

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid
Dirección General de Evaluación, Calidad e Innovación
CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Accenture. (2020). *Responsible AI: From Principles to Practice*. https://www.accenture.com/content/dam/accenture/final/a-com-migration/pdf/pdf-149/accenture-responsible-ai-final.pdfaccenture.com

Access Now. (2021). *The EU needs an Artificial Intelligence Act that protects fundamental rights.* https://www.accessnow.org/press-release/eu-artificial-intelligence-act-fundamental-rights/

Ada Lovelace Institute. (2022). *Rethinking data and rebalancing digital power*. https://www.adalovelaceinstitute.org/report/rethinking-data/adalovelaceinstitute.org+6adalovelaceinstitute.org+6adalovelaceinstitute.org+6

Administración Electrónica. (2025, 18 de febrero). *Cómo debe ser la gobernanza de datos en los modelos de IA open source.* https://administracionelectronica.gob.es/pae_Home/eu/pae_Actualidad/pae_Noticias/2025/Febrero/noticia-2025-02-18-Gobernanza-de-datos-modelos-IA-open-source.html

Administración Pública. (2020). *Gobernanza de datos como el camino hacia la inteligencia artificial.* https://www.administracionpublica.com/gobernanza-de-datos-como-el-camino-hacia-la-inteligencia-artificial/

Agencia Española de Protección de Datos (AEPD). (2020). *Guía para la gestión del riesgo y la evaluación de impacto en tratamientos de datos personales.* https://www.aepd.es/sites/default/files/2021-06/gestion-riesgo-y-evaluacion-impacto-en-tratamientos-datos-personales.pdfaepd.es

Agencia Española de Protección de Datos (AEPD). (2021). *Guía práctica para evaluaciones de impacto en protección de datos (DPIA).* https://www.aepd.es/guias/gestion-riesgo-y-evaluacion-impacto-en-tratamientos-datos-personales.pdfaepd.es

Agencia Española de Protección de Datos (AEPD). (2022). *La protección de datos en las relaciones laborales.* https://www.aepd.es/media/guias/guia-laboral.pdf

Agrawal, A., Gans, J., & Goldfarb, A. (2022a). From prediction to transformation. *Harvard Business Review,* 100–109. https://hbr.org/2022/11/from-prediction-to-transformation

AI CERTs. (2025). *AI+ Human Resources™ Certification*. https://store.aicerts.ai/certifications/business/ai-hr-certification/

AI Now Institute. (2019). *Algorithmic Accountability Policy Toolkit.* https://ainowinstitute.org/publications/algorithmic-accountability-policy-toolkit

AI Now Institute. (2020). *Algorithmic Accountability in the Workplace.* AI Now Institute.

AIHR Institute. (2025). *Empowering HR with AI: Champions of Diversity and Inclusion*. https://www.aihr-institute.com/blog/empowering-hr-with-ai-champions-of-diversity-and-inclusion

AIHR. (2023). *The state of AI in HR 2023*. https://www.aihr.com/blog/ai-in-hr-impact-adoption-automation/

AIHR. (2025). *Artificial Intelligence for HR Certificate Program.* https://www.aihr.com/courses/artificial-intelligence-for-hr-certification/

Ajunwa, I. (2020). The paradox of automation as anti-bias intervention. *Cardozo Law Review*, 41(5), 1671–1722.

AlgorithmWatch & Bertelsmann Stiftung. (2020). *Automating Society Report 2020*. https://automatingsociety.algorithmwatch.org/

AlgorithmWatch. (2021). *Algorithms and the world of work.* https://algorithmwatch.org/en/algorithms-and-the-world-of-work/

AlgorithmWatch. (2021). *Automating Society Report 2020.* https://automatingsociety.algorithmwatch.org/algorithmwatch.ch+2automatingsociety.algorithmwatch.org+2automatingsociety.algorithmwatch.org+2

AlgorithmWatch. (2023). *New study on AI in the workplace: Workers need control over algorithmic systems.* https://algorithmwatch.org/en/working-paper-ai-workplace-2023/

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bano, M., Zowghi, D., Gervasi, V., & Shams, R. (2023). AI for all: Operationalising diversity and inclusion requirements for AI systems. *arXiv preprint* arXiv:2311.14695. https://arxiv.org/abs/2311.14695

Bano, M., Zowghi, D., Mourao, F., Kaur, S., & Zhang, T. (2024). Diversity and Inclusion in AI for Recruitment: Lessons from Industry Workshop. *arXiv preprint* arXiv:2411.06066. https://arxiv.org/abs/2411.06066

Bar-Gil, O., Ron, T., & Czerniak, O. (2024). AI for the people? Embedding AI ethics in HR and people analytics projects. *Technology in Society*, 102527. https://doi.org/10.1016/j.techsoc.2024.102527

Barocas, S., & Selbst, A. D. (2016). Big Data's disparate impact. *California Law Review,* 104(3), 671–732. https://doi.org/10.2139/ssrn.2477899

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning*. https://fairmlbook.org

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and Machine Learning : Limitations and Opportunities.* The MIT Press.

Bassi, L. (2011). Raging debates in HR analytics. *People & Strategy,* 34(2), 14–18.

Batool, A., Zowghi, D., & Bano, M. (2024). AI governance: A systematic literature review. *AI and Ethics.* https://doi.org/10.1007/s43681-024-00653-w

Bellamy, R. K. E., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., Zhang, Y., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., & Mehta, S. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM *Journal of Research and Development*, 63(4–5). https://doi.org/10.1147/JRD.2019.2942287

Berocam. (2023). *Del Gobierno del Dato al Gobierno de la Inteligencia Artificial*. https://www.berocam.es/gobierno-del-dato-gobierno-de-la-ia/

Bersin, J. (2017, diciembre 19). *People analytics: Here with a vengeance*. Josh Bersin. https://joshbersin.com/2017/12/people-analytics-here-with-a-vengeance/

Bersin, J. (2018, junio 18). *AI in HR: A real killer app*. Josh Bersin. https://joshbersin.com/2018/06/ai-in-hr-a-real-killer-app/

Bersin, J. (2019). *HR technology market 2019: Disruption ahead*. Josh Bersin. https://joshbersin.com/hr-tech-disruptions-for-2019/

Bersin, J. (2022). *Understanding talent intelligence: A primer.* The Josh Bersin Company. https://joshbersin.com/understanding-talent-intelligence-a-primer/

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review,* 94(4), 991–1013. https://doi.org/10.1257/0002828042002561

Bielby, W. T. (2000). Minimizing workplace gender and racial bias. *Contemporary Sociology,* 29(1), 120–129. https://doi.org/10.2307/2654938

Bielby, W. T., & Baron, J. N. (1986). Men and women at work: Sex segregation and statistical discrimination. *American Journal of Sociology*, 91(4), 759–799. https://doi.org/10.1086/228350

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of the 1st *Conference on Fairness, Accountability and Transparency*, 81, 149–159. https://proceedings.mlr.press/v81/binns18a.html

Boatman, A. (2025). *AI courses for HR professionals: Top 11 recommendations.* AIHR. https://www.aihr.com/blog/ai-courses-for-hr-professionals/

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural*

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Information Processing Systems. https://arxiv.org/abs/1607.06520

Brey, P., & Dainow, · Brandt. (2023). Ethics by design for artificial intelligence. *AI and Ethics 2023* 4:4, 4(4), 1265–1277. https://doi.org/10.1007/S43681-023-00330-4

Brookings Institution. (2022). *AI and the Future of Work in Asia*. Brookings Institution.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* W. W. Norton & Company.

Buckland, R. (2022). AI, judges and judgement: Setting the scene (M-RCBG Associate Working Paper No. 220). *Harvard Kennedy School.* https://www.hks.harvard.edu/sites/default/files/centers/mrcbg/working.papers/Final_AWP_220.pdf

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness*, Accountability and Transparency, PMLR 81:77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

Business Wire. (2024, December 10). Most Leaders Have No AI Training. Business Wire. https://www.businesswire.com/news/home/20241210785102/en/Most-Leaders-Have-No-AI-Training-Which-Could-Negatively-Impact-Security-Privacy-and-Corporate-Competitiveness-Study-Finds

Calvard, T. S. (2021). Ethics of emotion recognition technology at work. *AI & Society,* 36(4), 1023–1032. https://doi.org/10.1007/s00146-020-01060-0

Cappelli, P. (2019). Your approach to hiring is all wrong. *Harvard Business Review,* 97(3), 48–58. https://hbr.org/2019/05/your-approach-to-hiring-is-all-wrong

Cedefop. (2021). *Digital and Green Skills: Defining and Measuring Competence*. https://www.cedefop.europa.eu/en/publications/3080

Chang, X. (2023). Gender Bias in Hiring: An Analysis of the Impact of Amazon's Recruiting Algorithm. Advances in Economics, *Management and Political Sciences*, 23(1), 134–140. https://doi.org/10.54254/2754-1169/23/20230367

Chartered Institute of Personnel and Development. (2020). *People analytics and the rise of HR data.* CIPD.

Cherry, M. A. (2016). *Beyond Misclassification: The Digital Transformation of Work.* Comparative Labor Law & Policy Journal. https://scholarship.law.slu.edu/cgi/viewcontent.cgi?article=1009&context=faculty

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data,* 5(2), 153–163. https://doi.org/10.1089/big.2016.0047

CIDOB. (2023). *La gobernanza de la inteligencia artificial: de solucionar los problemas técnicos a afrontar los retos éticos y políticos.* CIDOB. https://www.cidob.org/publicaciones/la-gobernanza-de-la-inteligencia-artificial-de-solucionar-los-problemas

CIPD (2023). *Communication and Trust in Digitally Enabled Workplaces*.

CIPD (2022). *People Analytics and Decision-Making*.

CIPD (2023). *People Analytics, Health Data and Bias*.

CIPD. (2021). *Learning and Skills at Work Survey 2021.* https://www.readkong.com/page/learning-and-skills-at-work-survey-2021-report-cipd-3664161

CIPD. (2021). *People analytics: Driving business performance with people data.*

Cole, M., Cant, C., Ustek Spilda, F., & Graham, M. (2022). Politics by Automatic Means? A Critique of Artificial Intelligence Ethics at Work. *Frontiers in Artificial Intelligence*, 5. https://doi.org/10.3389/frai.2022.869114

Comisión Europea. (2022). *Propuesta de Directiva sobre condiciones laborales en plataformas digitales*. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52021PC0762

Comisión Europea. (2023). *Directrices éticas para una IA fiable*. https://digital-strategy.ec.europa.eu/es/library/ethics-guidelines-trustworthy-ai

Consejo de la Unión Europea. (2000). Directiva 2000/78/CE del Consejo, de 27 de noviembre de 2000, por la que se establece un marco general para la igualdad de trato en el empleo y la ocupación. *Diario Oficial de las Comunidades Europeas,* L 303. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32000L0078

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & de Oliveira, N. (2022). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *arXiv*. https://arxiv.org/abs/2206.11922

Council of Europe. (2024). *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law.* https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

Coursera. (2024, octubre 30). *Enabling Generative AI Literacy: Skills Training for the Public Sector.* https://www.coursera.org/enterprise/articles/generative-ai-literacy-skills-training-public-sector-cm

Coursera. (2025). Generative *AI for Human Resources (HR) Professionals Specialization.* https://www.coursera.org/specializations/generative-ai-for-human-resources

Dankloff, M., Skoric, V., Sileno, G., Ghebreab, S., Van Ossenbruggen, J., & Beauxis-Aussalet, E. (2024). Analysing and organising human communications for AI fairness-related decisions: Use cases from the public sector. *arXiv*. https://arxiv.org/abs/2404.00022

Dastin, J. (2018). *Amazon scrapped a secret AI recruiting tool that showed bias against women.* Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

DataGuard. (2025). *AI anonymization: How tech supports data privacy in employee surveys.* https://www.dataguard.com/blog/ai-anonymization-employee-data/

Daugherty, P. R., & Wilson, H. J. (2018). Human + machine: Reimagining work in the age of AI. *Harvard Business Review Press.*

Davenport, T. H., Harris, J. G., & Shapiro, J. (2010). Competing on talent analytics. *Harvard Business Review*, 88(10), 52–58. https://hbr.org/2010/10/competing-on-talent-analytics

Day, M. (2023, October 25). *Half of Amazon's warehouse workers are injured after just 3 years.* Fortune. https://fortune.com/2023/10/25/amazon-worker-injuries-warehouse-study/

Daza, M. T., & Ilozumba, U. J. (2022). A survey of AI ethics in business literature: Maps and trends between 2000 and 2021. *Frontiers in Psychology*, 13, 8040. https://doi.org/10.3389/FPSYG.2022.1042661

Daza-Ramirez, M. T., & Luna Rizo, M. (2025). Automatización en el Aula: Desafíos Éticos de la IAG en la Educación Superior. In A. L. Iñiguez-Carrillo, M. T. Daza-Ramirez, & V. D. Arechiga-Cabrera (Eds.), *Inteligencia Artificial Generativa en la Educación Superior.* Tirant Lo Blanch.

De Stefano, V. (2022). Algorithmic management and collective labour rights: New challenges for labour standards and enforcement. *International Labour Review*, 161(2), 249–268. https://doi.org/10.1111/ilr.12209

De Stefano, V., & Taes, S. (2021). Algorithmic management and labour law: A cross-border overview. *ETUI Research Paper – Policy Brief* N.º 6/2021. https://www.etui.org/publications/algorithmic-management-and-labour-law-cross-border-overview

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

De Stefano, V., & Taes, S. (2022). Algorithmic management and collective bargaining. Transfer: *European Review of Labour and Research,* 28(4), 491–507. https://doi.org/10.1177/10242589221141055

Deloitte. (2022a). *Elevating humans in the digital workplace*. https://www.deloitte.com/global/en/services/consulting/analysis/elevating-humans-in-the-digital-workplace.htmlwww2.deloitte.com+2

Deloitte. (2022b). *Elevating the human experience*. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/about-deloitte/us-elevating-the-human-experience.pdf

Diligent Corporation. (2025). *AI governance & accountability: The foundation for responsible AI.* https://www.diligent.com/resources/blog/ai-governance

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. Journal of Personality and Social Psychology, 24(3), 285–290. https://doi.org/10.1037/h0033731

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world.* Basic Books.

Dong, Y., Ma, J., Wang, S., Chen, C., & Li, J. (2022). Fairness in Graph Mining: A Survey. IEEE Transactions on Knowledge and Data Engineering, 35(10), 10583–10602. https://doi.org/10.1109/TKDE.2023.3265598

Dutton, T. (2018). *The Global Landscape of AI Ethics Guidelines.*

Edelman Trust Institute. (2023). 2023 *Edelman Trust Barometer.* Edelman. https://www.edelman.com/sites/g/files/aatuss191/files/2023-01/2023%20Edelman%20Trust%20Barometer%20Global%20Report.pdf

Edwards, L., & Veale, M. (2018). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Information, Communication & Society,* 20(1), 1–18. https://doi.org/10.1080/1369118X.2017.1293125

EFFAT. (2023). *Recommendations for collective bargaining on algorithmic systems in the workplace*. https://effat.org/information/recommendations-for-collective-bargaining-on-algorithmic-systems-in-the-workplace/

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding explainability: Towards social transparency in AI systems. Proceedings of the 2021 CHI *Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3411764.3445188

El País. (2025, abril 9). *Adaptarse o morir: el reto de incorporar la IA en el trabajo.* https://elpais.com/proyecto-tendencias/2025-04-09/adaptarse-o-morir-el-reto-de-incorporar-la-ia-en-el-trabajo.htmlelpais.com+1elpais.com+1

España. (2015). Real Decreto Legislativo 2/2015, de 23 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto de los Trabajadores. *Boletín Oficial del Estado,* núm. 255. https://www.boe.es/buscar/act.php?id=BOE-A-2015-11430

España. (2018). Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. *Boletín Oficial del Estado*, núm. 294, 6 de diciembre de 2018. https://www.boe.es/eli/es/lo/2018/12/05/3/con

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. *St. Martin's Press.*

Eurofound. (2023a). *Human–robot interaction: What changes in the workplace?* Publications Office of the European Union. https://www.eurofound.europa.eu/en/publications/2024/human-robot-interaction-what-changes-workplace

Eurofound. (2023b). *Ethical digitalisation at work: From theory to practice.* Publications Office of the European Union. https://www.eurofound.europa.eu/en/publications/2023/ethical-digitalisation-work-theory-practice

European Agency for Safety and Health at Work. (2023). *AI-based systems for occupational safety and health.* Publications Office of the European Union. https://osha.europa.eu/en/publications/ai-based-systems-occupational-safety-and-health

European Commission, Joint Research Centre. (2020). *Robustness and explainability of artificial intelligence.* Publications Office of the European Union. https://ai-watch.ec.europa.eu/publications/robustness-and-explainability-artificial-intelligence_en

European Commission. (2020). *White paper on artificial intelligence: A European approach to excellence and trust* (COM (2020) 65 final). https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission. (2020). *White Paper on Artificial Intelligence: A European approach to excellence and trust.* https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission. (2021). *Artificial Intelligence Act: Article 10 – Data and Data Governance.* https://artificialintelligenceact.eu/article/10/

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act).* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

European Commission. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens*. https://op.europa.eu/en/publication-detail/-/publication/50c53c01-abeb-11ec-83e1-01aa75ed71a1/language-en

European Commission. (2023). *ESCO – European Skills, Competences, Qualifications and Occupations*. https://esco.ec.europa.eu/en/classification

European Commission. (2023). *Implementing AI governance: From framework to practice.* https://futurium.ec.europa.eu/en/european-ai-alliance/best-practices/implementing-ai-governance-framework-practice

European Commission. (2024). *Artificial Intelligence Act: Regulation (EU) 2024/1689*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

European Commission. (2025). *Article 4: AI literacy – EU Artificial Intelligence Act.* https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers

European Data Protection Board. (2020). *Directrices 05/2020 sobre el consentimiento con arreglo al Reglamento (UE) 2016/679*. https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_es.pdf

European Data Protection Board. (2021). *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default – Version Adopted After Public Consultation*. https://edpb.europa.eu

European Trade Union Confederation (ETUC). (2022). *AI in the workplace: Workers' rights and algorithmic management*. https://www.etuc.org/en/document/ai-workplace-workers-rights-and-algorithmic-management

European Trade Union Institute (ETUI). (2021). *Algorithmic management and collective bargaining.* https://www.etui.org/publications/algorithmic-management-and-collective-bargainingetui.org+4bibsonomy.org+4policycommons.net+4

European Trade Union Institute. (2022). *Exercising workers' rights in algorithmic management systems.* https://www.etui.org/publications/exercising-workers-rights-algorithmic-management-systems

European Union Agency for Fundamental Rights (FRA). (2020). *Getting the future right – Artificial intelligence and fundamental rights in the EU.* https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights

European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union*, L 119, 1–88. https://eurlex.europa.eu/eli/reg/2016/679/oj

Ferrara, E. (2023). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci 2024,* Vol. 6, Page 3, 6(1), 3. https://doi.org/10.3390/SCI6010003

Floridi, L. (2016). On Human Dignity as a Foundation for the Right to Privacy. *Philosophy and Technology*, 29(4), 307–312. https://doi.org/10.1007/S13347-016-0220-8/METRICS

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1), 1–15. https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707. https://doi.org/10.1007/S11023-018-9482-5

ForHumanity. (2023). Combatting Automation Bias. ForHumanity Inc. https://forhumanity.center/article/combatting-automation-bias/

Framework (AI RMF 1.0) Narayanan, A. (2018, February). Translation tutorial: 21 fairness definition and their politics. *Conference on Fairness, Account- Ability, and Transparency*. https://shubhamjain0594.github.io/post/tlds-arvind-fairness-definitions/

Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2016). On the (im)possibility of fairness. *arXiv preprint* arXiv:1609.07236. https://arxiv.org/abs/1609.07236

Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E. P., & Roth, D. (2019). A comparative study of fairness-enhancing interventions in machine learning. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 329–338. https://doi.org/10.1145/3287560.3287589

Frisby, K. (2025). AI compliance & safety for HR: Ensuring responsible and ethical AI copilots. LinkedIn. https://www.linkedin.com/pulse/ai-compliance-hr-ensuring-responsible-ethical-copilots-kelly-frisby-5hmif/

Fundación Cotec. (2022). *Marco de gobernanza algorítmica en las organizaciones*. https://cotec.es/informes/marco-de-gobernanza-algoritmica-en-las-organizaciones/

Fundación Cotec. (2023). *Buenas prácticas y retos de la IA en el empleo*. https://cotec.es/informes/buenas-practicas-y-retos-de-la-ia-en-el-empleo

Fundación Cotec. (2023). *Desafíos jurídicos de la inteligencia artificial en el empleo*. https://cotec.es/informes/desafios-juridicos-ia-empleo

Future of Life Institute. (2017). *Asilomar AI Principles*. https://futureoflife.org/open-letter/ai-principles/

Galarza, A. (2023, 22 de diciembre). *Revolutionizing performance reviews with generative AI.*

Gartner. (2023). *Gartner predicts 2024: AI's impact on the employee experience*. https://www.servicenow.com/lpebk/gartner-predicts-2024.html

Gartner. (2023). *Top Strategic HR Technology Trends.* Gartner.

Global Partnership on AI (GPAI). (2022). *AI and Work Project.* GPAI.

Gobierno de Navarra. (2025). *Navarra crea el Comité de Ética de la Oficina del Dato para garantizar la transparencia en el uso y desarrollo de la IA.* https://www.navarra.es/es/-/nota-prensa/navarra-crea-el-comite-de-etica-de-la-oficina-del-dato-para-garantizar-la-transparencia-en-el-uso-y-desarrollo-de-la-ia

Goldman Sachs. (2023). *Generative AI could raise global GDP by 7%.* https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percentgoldmansachs.com

Gray, M. L., & Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass.* Houghton Mifflin Harcourt.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review,* 102(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Guenole, N., Ferrar, J., & Feinzig, S. (2017). *The power of people: How successful organizations use workforce analytics to improve business performance.* FT Press.

Guenole, N., Ferrar, J., & Feinzig, S. (2017). *The Power of People: Learn How Successful Organizations Use Workforce Analytics to Improve Business Performance.* Pearson FT Press.

Gutelius, B., & Pinto, S. (2023). *Pain Points: Data on Work Intensity, Monitoring, and Health at Amazon Warehouses.* University of Illinois Chicago. https://doi.org/10.25417/UIC.24435124.V2

Güven, Ç., Alishahi, A., Brighton, H., Nápoles, G., Olier, J. S., Šafář, M., Postma, E., Shterionov, D., De Sisto, M., & Vanmassenhove, E. (2025). AI in support of diversity and inclusion. *arXiv preprint* arXiv:2501.09534. https://arxiv.org/abs/2501.09534

Hadley, E., Blatecky, A., & Comfort, M. (2024). Investigating algorithm review boards for organizational responsible AI governance. *AI and Ethics,* 5(1), 1–15. https://doi.org/10.1007/s43681-024-00574-8

Hardt, M., Price, E., & Srebro, N. (2016). *Equality of opportunity in supervised learning.* Advances in Neural Information Processing Systems. https://arxiv.org/abs/1610.02413

Harrington, C. (2024, 2 de enero). *AI may not steal your job, but it could stop you getting hired.* WIRED. https://www.wired.com/story/hilke-schellmann-algorithm-book-ai-jobs-hiring/

Harvard Business Review. (2020). *Reskilling in the Age of AI*. https://hbr.org/2020/10/reskilling-in-the-age-of-ai

Harvard Business Review. (2021). *How AI Is Changing Human Resources.* Harvard Business Review.

Harvard Business Review. (2021). *The New Skills Every Leader Needs*. https://hbr.org/2021/09/the-new-skills-every-leader-needs

Harvard Business Review. (2023). *How Generative AI Is Changing HR.* Harvard Business Review.

HealthBoxHR. (2025). *The role of AI in diversity, equity, and inclusion: Advancing HR strategies for a more inclusive workplace.* https://healthboxhr.com/blog/the-role-of-ai-in-diversity-equity-and-inclusion

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, 57(4), 657–674. https://doi.org/10.1111/0022-4537.00234

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Hendrycks, D. (2024). *Introduction to AI Safety, Ethics and Society.* Routledge.

High-Level Expert Group on AI (AI HLEG). (2019). *Ethics Guidelines for Trustworthy AI.* European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

High-Level Expert Group on AI (AI HLEG). (2019). *Policy and Investment Recommendations for Trustworthy AI.* https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligencedigital-strategy.ec.europa.eu

High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI.* European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

HireVue. (2025). *HireVue's AI explainability statement: An HR industry first*. https://www.hirevue.com/blog/hirevues-ai-explainability-statement-an-hr-industry-first

Hobsbawm, E. (1962). *Age of Revolution 1789-1848*. Vintage Books.

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudík, M., & Wallach, H. (2019). Improving fairness in machine learning systems: CHI Conference on Human Factors in Computing Systems (pp. 1–16). *ACM.* https://doi.org/10.1145/3290605.3300830

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019). What do industry practitioners need? Proceedings of the 2019 CHI *Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290605.3300830

HR Dive. (2023). *How to audit AI tools for bias*. https://www.hrdive.com/news/how-to-conduct-AI-bias-audit/687024/

HR Exchange Network. (2024). *Why artificial intelligence in HR needs governance*. https://www.hrexchangenetwork.com/dei/columns/why-artificial-intelligence-in-hr-needs-governance

HR Executive. (2025). *Announcing the AMS charter: A resource for ethical AI in HR.* https://hrexecutive.com/the-ethical-ai-blueprint-ams-charter-tackles-blind-spots-in-hr-policy/

HR Magazine. (2024). *AI and data privacy: How can HR remain compliant?* https://www.hrmagazine.co.uk/content/comment/ai-and-data-privacy-how-can-hr-remain-compliant

HR.com. (2023). *The ethical implications of using AI in HR.* https://www.hr.com/en/magazines/hr_strategy/april_2023_hr_strategy_planning_excellence/the-ethical-implications-of-using-ai-in-hr_lgnuju8q.html

Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50.

Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9

Hugill & Ip Solicitors. (2025). *Legal resolve: Practical tips for HR to balance AI, compliance and fairness.* https://www.hugillandip.com/2025/02/legal-resolve-practical-tips-for-hr-to-balance-ai-compliance-and-fairness/

Hunton Andrews Kurth LLP. (2024). *AI in HR: The current legal landscape*. https://www.hunton.com/insights/publications/ai-in-hr-the-current-legal-landscape

Hussain, A., & Hussain, A. (2025). Transparency and accountability: Unpacking the real problems of explainable AI. *AI & Society.* https://doi.org/10.1007/s00146-025-02302-0

IBM Institute for Business Value. (2020). Accelerating the journey to HR 3.0. https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/hr-3

IBM Institute for Business Value. (2022). *The Enterprise Guide to Closing the Skills Gap.* https://pact-for-skills.ec.europa.eu/document/download/ed4413e7-5d69-4b63-af5f-57f423d11a08_en?filename=The+enterprise+guide+to+closing+the+skills+gap.pdf

IBM Smarter Workforce Institute. (2017). *The business case for HR analytics*. IBM Corporation.

IBM. (2018). *Everyday ethics for artificial intelligence: A practical guide for designers & developers.* IBM Design for AI. https://www.ibm.com/design/ai/ethics/

IBM. (2020). *AI Fairness 360 Open Source Toolkit.* https://ai-fairness-360.org/ai-fairness-360.org+1new.pmean.com+1

IBM. (2023). Global executives and AI strategy for HR: How to tackle bias in algorithmic AI. https://www.ibm.com/think/insights/global-executives-and-ai-strategy-for-hr-how-to-tackle-bias-in-algorithmic-ai

IBM. (2023). *Responsible AI in HR: Guidelines for Implementation*. https://www.ibm.com/ai-human-resources

IBM. (2023a). *Responsible AI is a competitive advantage*. https://www.ibm.com/think/insights/responsible-ai-is-a-competitive-advantage

IBM. (2023b). *Responsible AI in communications: Building trust in customer interactions*. https://www.ibm.com/think/insights/responsible-ai-in-communications

IBM. (2023c). *Responsible AI in HR: Enhancing human resources with ethical AI*. https://www.ibm.com/think/insights/responsible-ai-in-hr

IBM. (2023d). *Responsible AI in HR administration: Streamlining processes ethically*. https://www.ibm.com/think/insights/responsible-ai-in-hr-administration

IBM. (2025). *IBM SkillsBuild*. https://skillsbuild.org

ICO. (n.d.). *AI and data protection risk toolkit. Information Commissioner's Office*. https://ico.org.uk/for-organisations/ai/

IEEE. (2020). *Ethically Aligned Design.* https://standards.ieee.org/industry-connections/ec/autonomous-systems/ead/

IEEE. (2022). *A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems.* IEEE Standards Association. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

Infobae. (2025, abril 10). *La Agencia de Protección de Datos española usará la Inteligencia Artificial para detectar riesgos que afecten a la privacidad*. https://www.infobae.com/espana/agencias/2025/04/10/la-agencia-de-proteccion-de-datos-usara-la-inteligencia-artificial-para-detectar-riesgos/

Information Commissioner's Office (ICO). (2022). *AI and data protection risk toolkit.* https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/ai-and-data-protection-risk-toolkit/

Insights for Professionals. (2021). *People analytics: Driving business performance with people data.* https://www.insightsforprofessionals.com/hr/leadership/people-analytics-driving-business-performance

International Labour Organization (ILO) & McKinsey & Company. (2021). *Building a Resilient Workforce.* https://www.ilo.org/global/publications/WCMS_776638/lang--en/index.htm

International Labour Organization (ILO). (2021). *AI and the world of work: Opportunities and risks.* https://www.ilo.org/global/topics/future-of-work/publications/WCMS_798346/lang--en/index.htm

International Labour Organization (ILO). (2021). *Mind the AI divide: Shaping a global perspective on the future of work*. https://www.ilo.org/publications/mind-ai-divide

International Labour Organization. (2022a). *Digital labour platforms and the future of work: Towards decent work in the online world.* https://www.ilo.org/publications/digital-labour-platforms-and-future-work-towards-decent-work-online-world

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

International Labour Organization. (2022b). EmergHigh-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI.* European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

International Labour Organization. (2023a). *Artificial intelligence and the future of work: A threat or a promise?* https://www.ilo.org/resource/other/artificial-intelligence-and-future-work-threat-or-promise

International Labour Organization. (2023b). *Negotiating the algorithm: Artificial intelligence and algorithmic management at work.* https://live.ilo.org/event/negotiating-algorithm-artificial-intelligence-and-algorithmic-management-work-2024-07-03

Jain, A., & Singhal, M. (2021). AI in HR: Revolutionizing human capital management. *International Journal of Business Research*, 21(2), 45–60.

Jain, A., & Singhal, M. (2021). Artificial intelligence in recruitment: A review. *Strategic HR Review,* 20(4), 123–129. https://doi.org/10.1108/SHR-03-2021-0023

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons,* 61(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence,* 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kahneman, D. (2011). *Thinking, Fast and Slow* [Book]. Penguin Books.

Kahneman, D., & Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Köchling, A., Wehner, M. C., & Wehner, M. (2023). Responsible artificial intelligence in human resources management: A review of the empirical literature. *AI and Ethics*. https://doi.org/10.1007/s43681-023-00325-1

Kondra, S., Medapati, S., Koripalli, M., Nandula, S. R. S. C., & Zink, J. (2025). AI and diversity, equity, and inclusion (DEI): Examining the potential for AI to mitigate bias and promote inclusive communication. *Journal of Artificial Intelligence and Machine Learning,* 3(1). https://jaim.sciforce.org/JAIM/article/view/249

Kourinian, A. (2024). Addressing transparency & explainability when using AI under global standards. *Bloomberg Law*. https://www.mayerbrown.com/-/media/files/perspectives-events/publications/2024/01/addressing-transparency-and-explainability-when-using-ai-under-global-standards.pdf

KPMG. (2023). *KPMG Global Tech Report 2023.* https://kpmg.com/es/es/informes-publicaciones/2023/11/kpmg-global-tech-report-2023.htmlservice.betterregulation.com+3kpmg.com+3kpmg.com+3

Kraus, M. W., & Keltner, D. (2009). Signs of socioeconomic status: A thin-slicing approach. *Psychological Science,* 20(1), 99–106. https://doi.org/10.1111/j.1467-9280.2008.02251.x

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Levenson, A. (2015). *Strategic analytics: Advancing strategy execution and organizational effectiveness.* Berrett-Koehler Publishers.

LinkedIn Learning. (2025). *Artificial Intelligence in Human Resources: Practical Applications*. https://www.linkedin.com/learning/ai-in-hr-practical-applications

LinkedIn Talent Solutions. (2024, 23 de abril). *AI in hiring: Balancing innovation and inclusion. LinkedIn*. https://www.linkedin.com/pulse/ai-recruitment-balancing-innovation-human-insight-hr4uwork-8lidc

LinkedIn. (2021). *Data is key to building responsible AI.* https://economicgraph.linkedin.com/blog/data-is-key-to-building-responsible-aieconomicgraph.linkedin.com

LinkedIn. (2024). *How do you apply governance when implementing AI in HR?* https://www.linkedin.com/pulse/how-do-you-apply-governance-when-implementing-ai-hr-seneena-x7xxf

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, 61(10), 36–43. https://doi.org/10.1145/3233231

Ma, S., Chen, Q., Wang, X., Zheng, C., Peng, Z., Yin, M., & Ma, X. (2024). Towards Human-AI Deliberation: Design and Evaluation of LLM-Empowered Deliberative AI for AI-Assisted Decision-Making. arXiv preprint arXiv:2403.16812 https://arxiv.org/abs/2403.16812

Macorva. (2024, 2 de octubre). *How AI reduces bias in employee performance reviews.* https://www.macorva.com/blog/how-ai-reduces-bias-in-employee-performance-reviews

Mansfield-Devine, S. (2015). The Ashley Madison affair. *Network Security*, 2015(9), 8–16. https://doi.org/10.1016/S1353-4858(15)30080-5

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Putting AI ethics into practice: The hourglass model of organizational AI governance. *arXiv preprint* arXiv:2206.00335. https://doi.org/10.48550/arXiv.2206.00335

Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *The International Journal of Human Resource Management,* 28(1), 3–26. https://doi.org/10.1080/09585192.2016.1244699

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology,* 6(3), 175–183. https://doi.org/10.1007/S10676-004-3422-1/METRICS

Mavrogiorgos, K., Kiourtis, A., Mavrogiorgou, A., Menychtas, A., & Kyriazis, D. (2024). Bias in Machine Learning: A Literature Review. *Applied Sciences 2024*, Vol. 14, Page 8860, 14(19), 8860. https://doi.org/10.3390/APP14198860

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine,* 27(4), 12–14. https://doi.org/10.1609/aimag.v27i4.1904

McKinsey & Company. (2022). *The state of AI in 2022—and a half decade in review.* https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review

McKinsey & Company. (2023). *The State of AI in 2023: Generative AI's Breakout Year.* McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year

Mehler, E. (2025). *AI and Inclusion: How Responsible AI Breaks Down Barriers.* LinkedIn. https://www.linkedin.com/pulse/ai-inclusion-how-responsible-breaks-down-barriers-eckhart-mehler-hrfqf/

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). *A survey on bias and fairness in machine learning.* ACM Computing Surveys, 54(6), 115:1–115:35. https://doi.org/10.1145/3457607

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Microsoft WorkLab. (2023a). *AI for HR: A Transformative Approach*. Microsoft. https://www.microsoft.com/en-us/microsoft-copilot/copilot-101/ai-for-hr

Microsoft WorkLab. (2023b). *The Next Great Disruption Is Hybrid Work—Are We Ready?* Microsoft. https://www.microsoft.com/en-us/worklab/work-trend-index/hybrid-work/

Minbaeva, D. (2020). Strategic HRM and data-driven decision making. *Human Resource Management Review*, 30(2), 100708.

Minbaeva, D. B. (2018). Building credible human capital analytics for organizational competitive advantage. *Human Resource Management,* 57(3), 701–713. https://doi.org/10.1002/hrm.21848

Ministerio de Trabajo y Economía Social (España). (2022). *Informe sobre digitalización y relaciones laborales.* https://www.mites.gob.es/ficheros/ministerio/analisis/boletines/2022/Informe_digitalizacion_relaciones_laborales_2022.pdf

Mitchell, T. M. (1997). *Machine learning.* McGraw-Hill.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21. https://doi.org/10.1177/2053951716679679

Mittelstadt, B. D., Allo, P., Taddeo,. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society*, 3(2). https://doi.org/10.1177/2053951716679679/ASSET/E715C393-1D70-4CE4-89D8-328D74B99E96/ASSETS/IMAGES/LARGE/10.1177_2053951716679679-FIG1.JPG

Mojsilović, A. (2018, septiembre 19). *Trust and transparency for AI on the IBM Cloud.* IBM Research Blog. https://newsroom.ibm.com/IBM-watson?item=30657

Mora-Cantallops, M., Sánchez-Alonso, S., García-Barriocanal, E., & Sicilia, M. A. (2021). Traceability for Trustworthy AI: A Review of Models and Tools. *Big Data and Cognitive Computing 2021*, Vol. 5, Page 20, 5(2), 20. https://doi.org/10.3390/BDCC5020020

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2021). From what to how: An initial review of publicly available AI ethics tools. *Science and Engineering Ethics,* 27(3), 1–31. https://doi.org/10.1007/s11948-019-00165-5

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences,* 109(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109

Mozilla Foundation. (2022). *Decolonizing AI & (Re)Defining a More Optimistic Future.* Mozilla Foundation. https://www.mozillafoundation.org/en/blog/decolonizing-ai-futures-at-mozfest/

Narayanan, A. (2018, February). Translation tutorial: 21 fairness definition and their politics. *Conference on Fairness, Account- Ability, and Transparency.* https://shubhamjain0594.github.io/post/tlds-arvind-fairness-definitions/

Nardinelli, C. (2007). Industrial Revolution and the Standard of Living. In D. R. Henderson (Ed.), The concise encyclopedia of economics (2nd ed.). *Liberty Fund.* https://www.econlib.org/library/Enc/IndustrialRevolutionandtheStandardofLiving.html

National Institute of Standards and Technology, U.S. Department of Commerce. NIST AI 100-1, *Artificial Intelligence Risk Management.* (2023). https://doi.org/10.6028/NIST.AI.100-1

Navascués y Cusí Abogados. (2025). *Los pilares fundamentales de la gobernanza en inteligencia artificial*. https://www.navascusi.com/gobernanza-inteligencia-artificial/

Nawrat, A. (2021, April 15). *HR tech gone wrong? Uber told to reinstate drivers after 'robo-firing.* UNLEASH. https://www.unleash.ai/hr-technology/court-rules-against-uber-robo-firing-employee-surveillance/

Nguyen, A. (2021). *The Constant Boss: Work Under Digital Surveillance.* https://datasociety.net/library/the-constant-boss/

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology,* 2(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nielsen Norman Group. (2023). *Usability Guidelines for Accessible Web Design.* Nielsen Norman Group https://media.nngroup.com/media/reports/free/Usability_Guidelines_for_Accessible_Web_Design.pdf

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35(4), 250–256. https://doi.org/10.1037/0022-3514.35.4.250

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism.* NYU Press. https://nyupress.org/9781479837243/algorithms-of-oppression/

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, 2(10), 100347. https://doi.org/10.1016/J.PATTER.2021.100347

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

OECD. (2019) *Principles for trustworthy AI*. https://oecd.ai/en/ai-principles

OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OECD. (2021). *AI and the Future of Skills,* Volume 1: Capabilities and Assessments. https://www.oecd.org/en/publications/ai-and-the-future-of-skills-volume-1_5ee71f34-en.html

OECD. (2021). *OECD Framework for the Classification of AI Systems*. https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htmoecd.org

OECD. (2021). *OECD principles on artificial intelligence*. https://www.oecd.org/going-digital/ai/principles/

OECD. (2021). *Recommendation of the Council on Artificial Intelligence*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OECD. (2021). *Using artificial intelligence in the workplace*. https://www.oecd.org/en/publications/using-artificial-intelligence-in-the-workplace_840a2d9f-en.html

OECD. (2021a). *OECD AI principles.* https://oecd.ai/en/ai-principles

OECD. (2021b). *OECD framework for the classification of AI systems*. https://oecd.ai/en/ai-principles

OECD. (2022). *OECD Employment Outlook 2022: Building Back More Inclusive Labour Markets.* OECD Publishing. https://doi.org/10.1787/1bb305a6-en

OECD. (2022). *Skills for a Digital World*. https://www.oecd.org/digital/skills-for-a-digital-world.pdf

OECD. (2023). *AI and the Future of Skills,* Volume 2. OECD Publishing.

OECD. (2023a). *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market.* OECD Publishing. https://doi.org/10.1787/08785bba-en

OECD. (2023b). *Using Artificial Intelligence in the Workplace: Opportunities, Risks and Policy Responses.* OECD Publishing. https://doi.org/10.1787/73d417f9-en

OECD. (2023c). The Impact of AI on the Workplace: Main Findings from the OECD AI Surveys of Employers and Workers. *OECD Social, Employment and Migration Working Papers*, No. 266. https://doi.org/10.1787/ea0a0fe1-en

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

ÓhÉigeartaigh, S. S., Whittlestone, J., Liu, Y., Zeng, Y., & Liu, Z. (2020). Overcoming barriers to cross-cultural cooperation in AI ethics and governance. *Philosophy & Technology,* 33(4), 571–593. https://doi.org/10.1007/s13347-020-00402-x

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown Publishing Group.

Optiblack. (2024). *Warden AI: Continuous bias auditing for HR tech.* https://optiblack.com/insights/ai-bias-audit-7-steps-to-detect-algorithmic-bias

Organisation for Economic Co-operation and Development. (2021). *AI and the future of skills, volume 1: Capabilities and assessments.* OECD Publishing. https://www.oecd.org/en/publications/ai-and-the-future-of-skills-volume-1_5ee71f34-en.html

Organisation for Economic Co-operation and Development. (2025). *Algorithmic management in the workplace.* OECD Publishing. https://www.oecd.org/en/publications/algorithmic-management-in-the-workplace_287c13c4-en.html

Organización Internacional del Trabajo (OIT). (2021). *El trabajo en la era digital: Políticas laborales y tecnológicas inclusivas*. https://www.ilo.org/global/publications/WCMS_798346/lang--es/index.htm

Organización Internacional del Trabajo (OIT). (2022). *Perspectivas sociales y del empleo en el mundo: Tendencias 2022.* https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms_848464.pdf

Organización Internacional del Trabajo. (2021). *Directrices sobre trabajo decente e inteligencia artificial*. https://www.ilo.org/artificial-intelligence

Osborne Clarke. (2024). *AI is reshaping HR and the workplace internationally and raising new legal issues*. https://www.osborneclarke.com/insights/ai-reshaping-hr-and-workplace-internationally-and-raising-new-legal-issues

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., Grieman, K., Havens, J. C., Jirotka, M., Kacorri, H., Karwowski, W., Kider, J., Konstan, J., Koon, S., Lopez-Gonzalez, M., Maifeld-Carucci, I., … Xu, W. (2023). Six Human-Centered Artificial Intelligence Grand Challenges. *International Journal of Human-Computer Interaction,* 39(3), 391–437. https://doi.org/10.1080/10447318.2022.2153320

Papagiannidis, E., Mikalef, P., & Conboy, K. (2024). Responsible artificial intelligence governance: A review and research agenda. *Information & Management*, 61(3), 103837. https://doi.org/10.1016/j.im.2024.103837

Papagiannidis, E., Mikalef, P., & Conboy, K. (2024). Responsible artificial intelligence governance: A review and research agenda. *Journal of Strategic Information Systems,* 34(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Parikh, R. B., Teeple, S., & Navathe, A. S. (2019). Addressing Bias in Artificial Intelligence in Health Care. *JAMA*, 322(24), 2377–2378. https://doi.org/10.1001/JAMA.2019.18058

Parlamento Europeo y Consejo de la Unión Europea. (2016). Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (Reglamento General de Protección de Datos). *Diario Oficial de la Unión Europea*, L 119, 1–88. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32016R0679

Paz Y Miño, G., & Khan, A. (2025, 17 de febrero). *Are AI-driven performance evaluations truly fair? People Matters Global.* https://me.peoplemattersglobal.com/article/hr-technology/the-ethics-of-ai-driven-performance-metrics-44318

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay,

É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research,* 12, 2825–2830. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Peretó Rovira, A. (2023). *Evolución e impacto de la inteligencia artificial desde una perspectiva económica: una mirada global, europea y regional* [Tesis doctoral, Universidad de Alicante]. RUA. https://rua.ua.es/dspace/handle/10045/136484

Plum. (2025). *Reducing bias in HR using artificial intelligence*. https://www.plum.io/report-reducing-bias-in-hr

Posthuma, R. A., & Campion, M. A. (2009). Age stereotypes in the workplace: Common stereotypes, moderators, and future research directions. *Journal of Management,* 35(1), 158–188. https://doi.org/10.1177/0149206308318617

Qin, C., Zhang, L., Cheng, Y., Zha, R., Shen, D., Zhang, Q., Chen, X., Sun, Y., Zhu, C., Zhu, H., & Xiong, H. (2023). *A Comprehensive Survey of Artificial Intelligence Techniques for Talent Analytics*. https://arxiv.org/abs/2307.03195v2

Qualee. (2025). *Leading with transparency: AI ethics and trust in the workplace*. https://www.qualee.com/blog/leading-with-transparency-ai-ethics-and-trust-in-the-workplace

Rachels, J. (1975). Why Privacy is Important. *Philosophy & Public Affairs*, 4(4), 323–333. https://www.jstor.org/stable/2265077

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2019). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *arXiv preprint,* arXiv:1906.09208. https://arxiv.org/abs/1906.09208

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. In Proceedings of the 2019 AAAI/ACM *Conference on AI, Ethics, and Society* (pp. 429–435). https://doi.org/10.1145/3306618.3314244

Randstad Research & CEOE. (2024). *Tendencias de RRHH 2024: Déficit de talento y adopción de IA.* https://www.ceoe.es/es/publicaciones/laboral/tendencias-de-rrhh-2024-deficit-de-talento-y-adopcion-de-iaceoe.es+1ceoe.es+1

Rasmussen, T., & Ulrich, D. (2015). Learning from practice: How HR analytics avoids being a management fad. *Organizational Dynamics,* 44(3), 236–242. https://doi.org/10.1016/j.orgdyn.2015.05.008

Redactor AI. (2025). *Data anonymization methods for AI-driven data governance*. https://redactor.ai/blog/data-anonymization-techniques-for-survey-data

Restack. (2024). *Governance playbook for AI systems.*

Rewari, V. (2024). *AI bias audit: 7 steps to detect algorithmic bias*. Optiblack. https://optiblack.com/insights/ai-bias-audit-7-steps-to-detect-algorithmic-bias

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International *Conference on Knowledge Discovery and Data Mining.* https://doi.org/10.1145/2939672.2939778

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International *Conference on Knowledge Discovery and Data Mining* (KDD 16) (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778

Rivera, L. A. (2012). Hiring as cultural matching: The case of elite professional service firms. *American Sociological Review*, 77(6), 999–1022. https://doi.org/10.1177/0003122412463213

Rodgers, W., & Al-Ali, M. (2022). An artificial intelligence algorithmic approach to ethical decision making in human resource management. *Journal of Business Research*, 139, 1050–1060. https://doi.org/10.1016/j.jbusres.2021.10.045

Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4.ª ed.). Pearson.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Sadhanand, J. (2025, 23 de enero). *How AI is driving diversity, equity, and inclusion* (DEI) in HR. LinkedIn. https://www.linkedin.com/pulse/how-ai-driving-diversity-equity-inclusion-dei-hr-joel-sadhanand-rxnmc

Samarawickrama, M. (2022). AI governance and ethics framework for sustainable AI and sustainability. *arXiv*. https://arxiv.org/abs/2210.08984

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. Paper presented at *ICA Data and Discrimination Preconference.* https://websites.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf

SAP (2021). *Ethics and AI at SAP*. https://www.sap.com/products/artificial-intelligence/ai-ethics.htmlsap.com

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., & Young, M. (2015). Hidden technical debt in machine learning systems. En *Advances in Neural Information Processing Systems* (Vol. 28, pp. 2503–2511). https://proceedings.neurips.cc/paper_files/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf

Sharot, T., et al. (2024). How human–AI feedback loops alter human perceptual, emotional, and social judgments. *Nature Human Behaviour*, 8(2), 235–244. https://www.nature.com/articles/s41562-024-02077-2

SHRM. (2025). *AI in HR: What you need to know—Data privacy and compliance*. https://www.shrm.org/topics-tools/topics/artificial-intelligence-in-the-workplace

Silver, C., & Lewins, A. (2014). *Using software in qualitative research* (2nd ed.). SAGE Publications.

Sison, A. J. G., Daza, M. T., Gozalo-Brizuela, R., & Garrido-Merchán, E. C. (2023). ChatGPT: More than a Weapon of Mass Deception, Ethical Challenges and Responses from the Human-Centered Artificial Intelligence (HCAI) Perspective. *International Journal of Human-Computer Interaction.* https://doi.org/10.1080/10447318.2023.2225931

Society for Human Resource Management (SHRM). (2024). *Ethical AI: Guidelines and best practices for HR professionals*. https://www.shrm.org/resourcesandtools/hr-topics/technology/pages/ethical-ai-guidelines.aspx

Society for Human Resource Management (SHRM). (2024). *Why diversity in AI makes better AI for all: The case for inclusive AI development*. https://www.shrm.org/topics-tools/flagships/ai-hi/why-diversity-in-ai-makes-better-ai-for-all--the-case-for-inclus

Society for Human Resource Management. (2023a). *AI and employee monitoring: Legal and ethical challenges.* https://www.shrm.org/topics-tools/topics/artificial-intelligence-in-the-workplace

Society for Human Resource Management. (2023b). *AI in HR: Enhancing efficiency without sacrificing empathy.* https://www.shrm.org/labs/resources/automate-hr-while-keeping-the-human-touch

Society for Human Resource Management. (2023c). *AI in labor relations: A double-edged sword*. https://www.shrm.org/topics-tools/employment-law-compliance/labor-relations

Society for Human Resource Management. (2023d). *Navigating employee communication in digital transitions.* https://www.shrm.org/labs/resources/building-a-connected-workforce-key-insights-on-employee-engagement

Springer, M. (2022). *Listening in Organizations: A Synthesis and Future Agenda. SSRN.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4058668

Springer, M. (2023). Transparency and explainability of AI systems: From ethical guidelines to practice. *AI and Ethics,* 3(1), 45–60. https://link.springer.com/article/10.1007/s43681-022-00123-4

Springer. (2024). *Special issue: Explainability in human-centric AI.* AI and Ethics. https://link.springer.com/collections/dbddddeceb

Stanmore School. (2025). *AI Ethics in HR: Professional Certificate for Ethical Practices.* https://www.stanmoreuk.org/Home/CourseDetail?courseId=57869

Statista. (2023). *Share of players in the NBA from 2010 to 2023, by ethnicity.* https://www.statista.com/statistics/1167867/nba-players-ethnicity/

Steele, C. M. (2010). *Whistling Vivaldi: How stereotypes affect us and what we can do.* W. W. Norton & Company.

Suresh, H., & Guttag, J. V. (2021). A framework for understanding unintended consequences of machine learning. *Communications of the ACM,* 64(3), 62–71. https://doi.org/10.1145/3430368

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15–42. https://doi.org/10.1177/0008125619867910

Tambe, P., Cappelli, P., & Yakubovich, V. (2023). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *Academy of Management Annals*, 17(1), 1–42.

Taori, R., & Hashimoto, T. B. (2022). Data feedback loops: Model-driven amplification of dataset biases. In Proceedings of the 39th International *Conference on Machine Learning* (pp. 33883–33920). https://proceedings.mlr.press/v202/taori23a/taori23a.pdf

Tarafdar, M., Cooper, C. L., & Stich, J.-F. (2019). The technostress trifecta – techno eustress, techno distress and design: Theoretical directions and an agenda for research. Information Systems Journal, 29(1), 6–42. https://doi.org/10.1111/isj.12169

Taylor Wessing. (2024). *The AI Act from an HR perspective*. https://www.taylorwessing.com/en/insights-and-events/insights/2024/09/the-ai-act-from-an-hr-perspective

Telefónica. (2020). *Principios de Inteligencia Artificial.* https://www.telefonica.com/es/wp-content/uploads/sites/4/2024/07/principios-inteligencia-artificial.pdflavanguardia.com+3telefonica.com+3telefonica.co+3

The Times. (2024, 20 de junio). *How to avoid the pitfalls when using AI to recruit new employees*. https://www.thetimes.co.uk/article/how-to-avoid-the-pitfalls-when-using-ai-to-recruit-new-employees-ml3vnlpgr

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25–29. https://doi.org/10.1037/h0071663

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry,* 16(10), 837–851. https://doi.org/10.1177/1077800410383121

Trades Union Congress (TUC). (2023). *Technology Managing People: The Worker Experience.* TUC. https://www.tuc.org.uk/research-analysis/reports/technology-managing-people-worker-experience

Trammell, E., & Korinek, A. (2023). *The Transformative Effects of AI on International Economics.* SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4658875papers.ssrn.com

Truby, J., Brown, R. D., Ibrahim, I. A., & Caudevilla Parellada, O. (2022). A sandbox approach to regulating high-risk artificial intelligence applications. *European Journal of Risk Regulation*, 13(2), 270–294. https://doi.org/10.1017/err.2022.19

Trustible. (2025). *AI Literacy Training*. https://www.trustible.ai/ai-literacy-training

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131.

UK Department for Work and Pensions. (2020). *Fraud and error in the benefit system: Financial year 2019 to 2020 estimates.* https://www.gov.uk/government/statistics/fraud-and-error-in-the-benefit-system-financial-year-2019-to-2020-estimatesassets.publishing.service.gov.uk+3gov.uk+3gov.uk+3

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

Ulrich, D., & Dulebohn, J. H. (2015). Are we there yet? What's next for HR? *Human Resource Management Review,* 25(2), 188–204. https://doi.org/10.1016/j.hrmr.2015.01.004

Unesco (2021). *Competency Frameworks in Education and Employment*. https://unesdoc.unesco.org/ark:/48223/pf0000377250

Unesco (2021). *Recommendation on the Ethics of Artificial Intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNI Global Union. (2020). *Top 10 Principles for Ethical AI.* https://uniglobalunion.org/news/uni-global-union-releases-top-10-principles-for-ethical-ai/

UNI Global Union. (2021). *The Amazon Panopticon.* https://uniglobalunion.org/wp-content/uploads/amazon_panopticon_en_final.pdfuniglobalunion.org+2uniglobalunion.org+2uniglobalunion.org+2

UNI. (2021). *UNE-EN ISO 30414: Human capital reporting.*

United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. https://sdgs.un.org/2030agenda

Université de Montréal. (2024) *Rapport de recherche – Déclaration de Montréal et notion de responsabilité.* https://declarationmontreal-iaresponsable.com/

Upadhyay, A. K., & Khandelwal, K. (2018). Applying artificial intelligence in human resource management: Opportunities and challenges. *Journal of Strategic Human Resource Management,* 7(3), 15–22.

Upadhyay, A. K., & Khandelwal, K. (2018). Applying artificial intelligence: Implications for recruitment. *Strategic HR Review*, 17(5), 255–258. https://doi.org/10.1108/SHR-07-2018-0051

Ustun, B., Spangher, A., & Liu, Y. (2019). Actionable recourse in linear classification. In Proceedings of the 2019 AAAI/ACM *Conference on AI, Ethics, and Society* (FAT)* (pp. 10–19). https://doi.org/10.1145/3287560.3287566

Van der Merwe, M. (2025, 28 de marzo). *AI in talent management: Use and limitations in HR*. AIHR. https://www.aihr.com/leading-hr/ai-in-talent-management/

Varshney, K. R. (2022). *Engineering trustworthy AI.* http://www.trustworthymachinelearning.com/trustworthymachinelearning.pdf

Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2), 1–17. https://doi.org/10.1177/2053951717743530

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97–112. https://doi.org/10.9785/cri-2021-220402

Veale, M., & Edwards, L. (2018). Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling. *Computer Law & Security Revie*w, 34(2), 398–404. https://doi.org/10.1016/j.clsr.2017.12.002

Verma, S., & Rubin, J. (2018). Fairness definitions explained. Proceedings - *International Conference on Software Engineering,* 1–7. https://doi.org/10.1145/3194770.3194776

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law,* 7(2), 76–99. https://doi.org/10.1093/idpl/ipx005

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law,* 7(2), 76–99. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841–887. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3063289

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review,* 41, 105567. https://doi.org/10.1016/j.clsr.2021.105567

Warden AI. (s.f.). *Warden AI.* https://www.warden-ai.com/

Wexler, J., Pushkarna, M., Bolukbasi, T., Wattenberg, M., Viégas, F., & Wilson, J. (2019). The What-If Tool: Interactive probing of machine learning models. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 56–65. https://doi.org/10.1109/TVCG.2019.2934619

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. In Proceedings of the 2019 AAAI/ACM *Conference on AI, Ethics, and Society* (pp. 195–200). https://doi.org/10.1145/3306618.3314289

Williams, J. C., & Dempsey, R. (2014). *What works for women at work: Four patterns working women need to know*. NYU Press.

Williams, J. C., Phillips, K. W., & Hall, E. V. (2016). *Double jeopardy? Gender bias against women of color in science.* Center for WorkLife Law, UC Hastings College of the Law.

Wilson, C., Hoffman, M., & Morgenstern, J. (2021). Building and auditing fair algorithms: A case study in candidate screening. In Proceedings of the 2021 ACM *Conference on Fairness, Accountability, and Transparency* (FAccT '21) (pp. 666–681). https://doi.org/10.1145/3442188.3445928

Winkler, T. (2023). Artificial intelligence explainability requirements of the AI Act and metrics for measuring compliance. In *Lecture Notes in Information Systems and Organisation* (Vol. 75). Springer. https://link.springer.com/chapter/10.1007/978-3-031-80122-8_8

Winkler, T. (2024). Artificial intelligence explainability requirements of the AI Act and metrics for measuring compliance. In *Lecture Notes in Information Systems and Organisation* (Vol. 75). Springer. https://doi.org/10.1007/978-3-031-80122-8_8

World Economic Forum. (2020). *The future of jobs report 2020.* https://www.weforum.org/publications/the-future-of-jobs-report-2020/

World Economic Forum. (2021) *Toolkit Human-Centred AI for HR.* https://www.weforum.org/publications/human-centred-ai-for-hr-state-of-play-and-the-path-ahead/toolkit/

World Economic Forum. (2021). *Human-centred AI for HR: State of play and the path ahead.* https://www.weforum.org/publications/human-centred-ai-for-hr-state-of-play-and-the-path-ahead/

World Economic Forum. (2022). *HR 4.0: Shaping People Strategies in the Fourth Industrial Revolution*. https://www.weforum.org/publications/hr4-0-shaping-people-strategies-in-the-fourth-industrial-revolution/

World Economic Forum. (2023). *Global AI adoption index.* https://www.weforum.org/publications/global-ai-adoption-index-2023/

World Economic Forum. (2023). *Jobs of tomorrow: Large language models and jobs.* https://www.weforum.org/publications/jobs-of-tomorrow-large-language-models-and-jobs/

World Economic Forum. (2023). *The Future of Jobs Report 2023.* https://www.weforum.org/publications/the-future-of-jobs-report-2023/

World Economic Forum. (2023a). *Empowering workers in the age of AI.* https://www.weforum.org/stories/2023/12/empowering-tomorrow-s-workforce-in-the-ai-age/

World Economic Forum. (2024a). *Building resilient and healthy workplaces: A report on mental health at work.* https://www3.weforum.org/docs/WEF_Building_Resilient_and_Healthy_Workplaces_2024.pdf

World Economic Forum. (2024b). *Putting skills first: Opportunities for building efficient and equitable labour markets.* https://www3.weforum.org/docs/WEF_Putting_Skills_First_2024.pdf

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE DERECHOS SOCIALES Y AGENDA 2030
Plan de Recuperación, Transformación y Resiliencia
Comunidad de Madrid Dirección General de Evaluación, Calidad e Innovación CONSEJERÍA DE FAMILIA, JUVENTUD Y ASUNTOS SOCIALES

World Economic Forum. (2024c). *AI for impact: A playbook for social innovation.* https://www3.weforum.org/docs/WEF_AI_for_Impact_Social_Innovation_2024.pdf

Zafar, M. B., Valera, I., Gomez Rodriguez, M., & Gummadi, K. P. (2017). Fairness constraints: Mechanisms for fair classification. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017). *Proceedings of Machine Learning Research,* 54, 962–970. https://arxiv.org/abs/1507.05259

Zarsky, T. Z. (2017). Incompatible: The GDPR in the Age of Big Data. *Seton Hall Law Review,* 47(4), 995–1020. https://scholarship.shu.edu/shlr/vol47/iss4/2/

Zhang, A., Walker, O., Nguyen, K., Dai, J., Chen, A., & Lee, M. K. (2023). Deliberating with AI: Improving decision-making for the future through participatory AI design and stakeholder deliberation. *Proceedings of the ACM on Human-Computer Interaction,* 7(CSCW1), 1–32. https://doi.org/10.1145/3579471

Zliobaite, I. (2017). Measuring discrimination in algorithmic decision making. *Data Mining and Knowledge Discovery,* 31(4), 1060–1089. https://doi.org/10.1007/s10618-017-0506-1

Primera edición: julio 2025
ISBN: 978-84-09-73740-6
Copyright © **2025, Comunicación de Valor Añadido S.L.U.**
Información y contacto: Marisa Cruzado, coordinadora de IA+Igual.
mcruzado@cvalora.com
Coaturores: José Luis Poveda, Efrén Díaz, Marco Tulio Daza y Ana Cantalapiedra (DATAI - Universidad de Navarra)

Aviso legal y reserva de derechos
Todos los derechos están reservados. Queda estrictamente prohibida, sin autorización escrita del titular del copyright, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, incluidos la reprografía y el tratamiento informático, conforme a las sanciones establecidas en el ordenamiento jurídico vigente. Los contenidos de esta publicación son responsabilidad exclusiva del titular del copyright, por lo que la administración competente no asume responsabilidad alguna sobre dichos contenidos. Para solicitar la autorización de su reproducción total o parcial escribir al email: cva@cvalora.com



Financiado por la Unión Europea
NextGenerationEU

GOBIERNO
DE ESPAÑA

MINISTERIO
DE DERECHOS SOCIALES
Y AGENDA 2030

Plan de Recuperación,
Transformación
y Resiliencia

Comunidad de Madrid

Dirección General
de Evaluación, Calidad e Innovación
CONSEJERÍA DE FAMILIA,
JUVENTUD Y ASUNTOS SOCIALES