

I

ESTUDIOS DE TEORÍA Y FILOSOFÍA DEL DERECHO

Desafíos y oportunidades de los planes de igualdad en la era de la inteligencia artificial (IA)¹

Challenges and opportunities of equality plans in the age of artificial intelligence (AI)

Por M.^a CRUZ DÍAZ DE TERÁN VELASCO
Universidad de Navarra

RESUMEN

Este artículo analiza algunos de los desafíos que plantea el uso de la inteligencia artificial (IA) en el ámbito laboral desde una perspectiva ético-jurídica. A partir del concepto de sesgo y su impacto en los sistemas automatizados, se examinan casos concretos de discriminación hacia las mujeres. Se propone integrar medidas específicas en los Planes de Igualdad, tanto de transparencia como de formación, que permitan supervisar el funcionamiento de los algoritmos y prevenir desigualdades. La aportación de la Filosofía del Derecho permite enmarcar este análisis desde una perspectiva crítica. Hacerlo así refuerza la necesidad de una gobernanza ética de la IA que sitúe la libertad, la igualdad y la protección de los derechos fundamentales en el centro del desarrollo normativo.

Palabras clave: Inteligencia artificial (IA), sesgos de género, Plan de Igualdad, gobernanza ética.

¹ Este trabajo se ha realizado en el marco de la Red de formación y fortalecimiento del impacto femenino en los ecosistemas de innovación (620RT0013-Red-Winn), liderado por M.^a Cruz Díaz de Terán y financiado por el Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (Programa Cyted).

ABSTRACT

This article analyzes some of the challenges posed by the use of artificial intelligence (AI) in the workplace from a legal-ethical perspective. Based on the concept of bias and its impact on automated systems, specific cases of discrimination against women are examined. It is proposed to integrate specific measures in the Equality Plans, both in terms of transparency and training, to monitor the operation of algorithms and prevent inequalities. The contribution of the Philosophy of Law makes it possible to frame this analysis from a critical perspective. Doing so reinforces the need for ethical governance of AI that places freedom, equality and the protection of fundamental rights at the center of regulatory development.

Keywords: Artificial intelligence (AI), gender bias, Equality plans, ethical governance.

SUMARIO: 1. INTRODUCCIÓN. – 2. NOCIONES FUNDAMENTALES. 2.1 *Conceptos relacionados con la Inteligencia Artificial (IA).* 2.2 *Los sesgos de género en la IA.* 2.2.1 *Definición de sesgo.* 2.2.2 *Tipos de sesgos.* 2.2.3 *Casos de sesgos de género en aplicaciones de IA en el ámbito laboral.* – 3. *LOS PLANES DE IGUALDAD.* 3.1 *Medidas de transparencia e igualdad en la creación y evaluación de los algoritmos.* 3.2 *Medidas de formación en igualdad.* – 4. *ÉTICA Y RESPONSABILIDAD EN EL USO DE HERRAMIENTAS DE IA.* 4.1 *Principios ético-jurídicos fundamentales.* 4.2 *Filosofía del Derecho, Ética y sesgos en la IA.* – 5. *CONCLUSIONES.*

SUMMARY: 1. INTRODUCTION. – 2. FUNDAMENTAL NOTIONS. 2.1 *Concepts related to Artificial Intelligence (AI).* 2.2 *Gender bias in AI.* 2.2.1 *Definition of bias.* 2.2.2 *Types of bias.* 2.2.3 *Cases of gender bias in AI applications in the workplace.* – 3. *EQUALITY PLANS.* 3.1 *Transparency and equality measures in the creation and evaluation of algorithms.* 3.2 *Equality training measures.* – 4. *ETHICS AND RESPONSIBILITY IN THE USE OF AI TOOLS.* 4.1 *Fundamental ethical-legal principles.* 4.2 *Philosophy of Law, Ethics and bias in AI.* – 5. *CONCLUSIONS.*

1. INTRODUCCIÓN

La inteligencia artificial (IA) se ha consolidado como una realidad dispuesta a romper las reglas establecidas en ámbitos clave como el empleo, el mercado, la educación o la salud. Su integración, cada vez mayor en la vida cotidiana, plantea desafíos que requieren una reflexión serena y rigurosa. Entre las disciplinas llamadas a responder

a estos retos, el Derecho ocupa un lugar central y, en este contexto jurídico, la Filosofía del Derecho no puede permanecer al margen. La Filosofía del Derecho busca precisamente iluminar el sentido y los límites del Derecho ante los nuevos desafíos que se plantean. En este ámbito su función, como ha firmado Pérez Luño, es ser «la conciencia de la IA, que es inteligencia sin conciencia»².

Una de las cuestiones más relevantes en este marco es el análisis de los sesgos en los algoritmos, en particular, los sesgos de género. Y es que la reproducción de estos sesgos puede incidir de manera determinante en la expansión de desigualdades históricas, o, por lo contrario, en su corrección si se detectan y evitan³. Con otras palabras, los algoritmos que orientan decisiones en estos ámbitos tienen el potencial de consolidar patrones discriminatorios o, por el contrario, pueden convertirse en instrumentos que promuevan una mayor igualdad. La diferencia entre una y otra posibilidad radica, en gran medida, en la calidad y representatividad de los datos con los que son entrenados. De ahí la importancia de un enfoque jurídico crítico en el desarrollo y aplicación de la IA; un enfoque que no solo identifique estos sesgos, sino que proponga soluciones efectivas para evitar que la tecnología amplíe las desigualdades en lugar de contribuir a su desaparición.

En España, los Planes de Igualdad son instrumentos estratégicos destinados a promover la igualdad efectiva entre mujeres y hombres en los ámbitos laboral, institucional y social. Su propósito es identificar y erradicar las desigualdades de género existentes mediante medidas concretas que garanticen la igualdad de oportunidades, la no discriminación y la plena participación de mujeres y hombres en todos los aspectos de la vida⁴. En este contexto, los Planes de Igualdad pueden desempeñar un papel clave en la lucha contra los sesgos de género, contribuyendo a reducir las desigualdades.

Y es que la integración de herramientas de inteligencia artificial en los Planes de Igualdad es un fenómeno en expansión. Estas tecnologías ofrecen la posibilidad de analizar, detectar y corregir desigualdades entre mujeres y hombres en coherencia con los objetivos de estos planes. Sin embargo, como se ha señalado, también conllevan el riesgo de reproducir o amplificar los sesgos preexistentes. Ante esta realidad, resulta imprescindible reflexionar sobre qué mecanismos deben incorporarse en los Planes de Igualdad para garantizar que los algoritmos sean diseñados y aplicados sin perpetuar discriminaciones.

² Cfr. PÉREZ LUÑO, Antonio Enrique; «La Filosofía del Derecho y sus tareas: presente y futuro», *Anuario de Filosofía del Derecho*, vol. 39, 2024, p. 234.

³ MENDONÇA DE LIMA, Rosileine, R, PISKER, BARBARA AND CORRÊA, VICTOR Silva; «Gender Bias in Artificial Intelligence», *Australian Journal of Telecommunications and the Digital Economy*, 11/2, 2023, pp. 8-30.

⁴ Real Decreto 901/2020, de 13 de octubre, por el que se regulan los Planes de Igualdad y su registro.

Esta cuestión, a su vez, plantea un desafío más amplio: ¿pueden los Planes de igualdad no solo prevenir sesgos en la IA, sino también orientar su uso ético? En otras palabras, ¿es posible aprovechar estas tecnologías como herramientas para avanzar en la igualdad de género, evitando que refuercen estereotipos o prácticas discriminatorias?

El presente trabajo tiene como propósito abordar estos interrogantes. Para ello, en primer lugar se delimitarán los conceptos clave vinculados a la IA, los sesgos y los Planes de Igualdad. Posteriormente, se analizará si es posible que los Planes pueden ser un componente clave en la regulación y supervisión de la IA. Por último, se hará una serie de recomendaciones ético-jurídicas dirigidas a asegurar que la incorporación de la IA en los Planes de Igualdad contribuya, de manera efectiva, a la promoción de la igualdad entre mujeres y hombres.

2. NOCIONES FUNDAMENTALES

2.1 Conceptos relacionados con la Inteligencia Artificial (IA)

La IA es una disciplina científica dedicada al desarrollo de sistemas informáticos que realizan tareas habitualmente asociadas a la mente humana, como el aprendizaje, el razonamiento lógico⁵. A través de algoritmos y modelos matemáticos, procesa grandes volúmenes de datos para identificar patrones y generar decisiones automáticas basadas en ellos. Si bien la inteligencia artificial no es una tecnología nueva, los avances recientes en el procesamiento de datos han favorecido su expansión y aplicación en distintos ámbitos. Hoy en día, las máquinas pueden razonar, aprender y comprender con una rapidez y eficiencia que supera, en muchos casos, la capacidad humana. Estas tecnologías permiten analizar grandes volúmenes de información en muy poco tiempo y extraer conocimientos útiles para una amplia variedad de sectores⁶.

La IA se basa en el uso de algoritmos. Los algoritmos son secuencias de operaciones ordenadas y finitas diseñadas para resolver un problema específico⁷. En el contexto de la programación, un

⁵ Real academia española; *Diccionario de la lengua española*, 23.^a ed., <https://dle.rae.es/inteligencia#2DxmhCT>.

⁶ A día de hoy, las aplicaciones de la inteligencia artificial abarcan un amplio abanico de ámbitos y funciones, que van desde tareas cotidianas hasta procesos complejos y especializados. Por poner algunos ejemplos: asistentes virtuales (Siri, Alexa), legales (Sof-IA, Conversa); sanitarios (IMPAI, Syntomate) o traductores automáticos (Google Translate, DeepL) son solo una muestra de su creciente implantación. Sobre este tema, *vid*, entre otros, KARMAKAR, Pragyna, SINHA, Satarupa & PAL, Debrupa; «Artificial Intelligence», *International Journal of Advanced Research in Science, Communication and Technology*, vol. 4/2, 2024, pp. 79-87.

⁷ Real academia española; *Diccionario*, *op. cit.* <https://dle.rae.es/algoritmo#1nmLTsh>.

algoritmo es un conjunto de instrucciones ordenadas y precisas que guían el proceso de resolución de un problema o la realización de una tarea. Es el paso previo a la creación del código, donde se define el procedimiento necesario para alcanzar una solución. Un programa informático es, en esencia, un conjunto de algoritmos codificados en un lenguaje de programación, listos para ser ejecutados por un ordenador⁸.

Una de las principales ramas de la IA es el aprendizaje automático (o *Machine Learning*). El aprendizaje automático permite que una máquina aprenda de los datos, reconociendo patrones y realizando predicciones, incluso para escenarios que no se habían previstos inicialmente. Este enfoque se basa en la gestión de datos, la evaluación y la ejecución de modelos. El proceso incluye, en primer lugar, la recolección, filtrado y validación de datos para el entrenamiento y prueba del modelo; en segundo lugar, la configuración de los objetivos y enfoques del modelo, la selección de variables y algoritmos; en tercer lugar, la evaluación de la precisión del modelo mediante una verificación de sus resultados. Por último, el modelo se despliega para trabajar con nuevos datos. Existen tres variantes principales de aprendizaje automático: supervisado, no supervisado y por refuerzo⁹.

En el aprendizaje supervisado, el modelo se entrena con ejemplos etiquetados y se ajusta para predecir o clasificar nuevos datos basándose en patrones identificados, es decir, ejemplos en los que tanto las entradas como las salidas (resultados) ya son conocidas. El objetivo es que el modelo «aprenda» a predecir o clasificar correctamente nuevas entradas basándose en lo que ha aprendido de los ejemplos previos¹⁰. En el aprendizaje no supervisado, el modelo trabaja con datos no etiquetados, es decir, solo tiene acceso a las entradas, pero no a las salidas o resultados. El objetivo es encontrar patrones, relaciones o estructuras en los datos, como grupos o tendencias¹¹. Por último, el aprendizaje por refuerzo implica que un agente (modelo) interactúa con un entorno, tomando acciones y recibiendo recompensas o penalizaciones según el resultado de esas acciones. El objetivo es que el agente aprenda a maximizar su recompensa a lo largo del

⁸ BOGDANOVIC, Zorica, «Artificial Intelligence in Federal information Processing Systems», *American Journal of Computer Science and Information Technology*, vol. 9/7:99, 2021, p. 1.

⁹ KAUR, Gagandeep, KUMAR, Santosh; «Applications of Machine Learning in Diverse Sector: A Study», *International Journal of Contemporary Technology*, vol. 5/2 (2023) pp. 344-345.

¹⁰ Si se dispone de un conjunto de datos con imágenes de animales y sus etiquetas (por ejemplo, «gato», «perro»), el modelo aprende a identificar nuevas imágenes y clasificarlas correctamente según esas etiquetas.

¹¹ Si se dispone de un conjunto de datos de clientes con información como edad, ingresos y preferencias, el modelo podría agrupar a los clientes en segmentos similares (por ejemplo, «clientes jóvenes con alto ingreso» o «clientes mayores con bajo ingreso»), sin saber de antemano cuáles son esos grupos.

tiempo, mejorando su toma de decisiones en función de la retroalimentación obtenida¹².

Por tanto, en cualquiera de las tres variantes, el aprendizaje automático permite que una máquina aprenda y mejore a medida que procesa más datos¹³.

2.2 Los sesgos de género en la IA

Dado que los modelos de IA aprenden a partir de los datos con los que se entrenan, cualquier sesgo presente en esos datos tiene el potencial de replicarse, e incluso ampliarse, por los algoritmos. Este fenómeno puede generar riesgos de discriminación, especialmente en contextos sensibles como el sexo, la etnia o la clase social. Por lo tanto, la identificación y corrección de estos sesgos es el primer paso imprescindible para garantizar que los sistemas de IA fomenten la igualdad, y no perpetúen o aumenten las desigualdades ya existentes.

2.2.1 Definición de sesgo

El Parlamento Europeo define el sesgo como «toda percepción personal o social prejuiciosa de una persona o grupo sobre la base de sus características personales»¹⁴. En términos técnicos, se trata de una desviación sistemática respecto a un estándar o una media estadística¹⁵. En el contexto de la inteligencia artificial (IA), hablamos de sesgo cuando una tecnología, debido a un tratamiento desigual o prejuicioso de los datos, no ofrece resultados adecuados para ciertos grupos¹⁶.

Los sesgos en la IA pueden tener dos orígenes: los datos y el diseño del modelo. En el caso de los datos, el problema aparece cuando determinados grupos están «infra o sobrerrepresentados». Esto genera resultados distorsionados al aplicar el modelo. Si entendemos la IA como una tecnología de inferencia estadística, sus resultados dependerán, de forma directa, de la calidad de los datos utilizados en su entrenamiento. A pesar de ello, la detección de estos sesgos no siempre es sencilla. Además está el hecho de que, en muchos casos, se

¹² Este sería el caso de un robot que aprende a moverse por un laberinto. El robot recibe una recompensa cuando llega al final del laberinto y penalizaciones si toma decisiones equivocadas. A través de ensayo y error, aprende a encontrar la mejor ruta.

¹³ MARTÍN LÓPEZ, Jorge; «Inteligencia artificial, sesgos y no discriminación en el ámbito de la inspección tributaria», *Crónica tributaria*, vol. 182, 2022, p. 58.

¹⁴ European Parliament, *Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies* [2020/2012(INL)], artículo 4.1).

¹⁵ DANKS, David y LONDON, Alex John, «Algorithmic Bias in Autonomous Systems», Conference: Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), 2017, p. 4692.

¹⁶ European Parliament, *Resolution of 20 October 2020...*, *op. cit.*, artículo 4.

desconoce su existencia. Junto con ello, el llamado «efecto eficacia» puede dificultar su identificación: si el sistema ofrece resultados que parecen satisfactorios, se tiende a asumir que no es necesario revisarlo ni corregirlo¹⁷.

Los sesgos pueden surgir también directamente de los algoritmos, donde las decisiones sobre qué variables priorizar durante el proceso de aprendizaje pueden inducir errores sistemáticos que favorecen a ciertos grupos en detrimento de otros. Estos sesgos estadísticos provienen de la selección de parámetros que, de manera constante, desfavorecen a determinados colectivos en comparación con la media estadística. Un riesgo añadido es que estos sesgos se mantengan a lo largo del tiempo, dado que los modelos se reentrenan utilizando datos que perpetúan las mismas debilidades¹⁸.

2.2.2 Tipos de sesgos

El sesgo en IA puede manifestarse de diversas formas, como de género, étnico o económico. Y lo habitual es que afecte a grupos insuficientemente representados en los datos utilizados para entrenar modelos de aprendizaje automático. De hecho, en términos informáticos, el concepto de sesgo algorítmico surgió de la observación de tecnologías de reconocimiento facial y vigilancia predictiva, las cuales reproducían desigualdades existentes bajo la vigilancia dirigida por los Estados¹⁹. A su vez, estos sesgos pueden estar presentes en diversos tipos de datos, como números, textos, imágenes o sonidos.

Entre los sesgos más comunes que impactan en los colectivos vulnerables, se encuentran los siguientes²⁰:

a) Sesgo de selección de datos: como se ha indicado anteriormente, este sesgo se produce cuando los datos utilizados para entrenar al algoritmo no representan de manera adecuada a los grupos más vul-

¹⁷ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Building Trust in Human-Centric Artificial Intelligence* [COM (2019) 168 final], p. 6; European Commission, *White Paper on Artificial Intelligence – A European approach to excellence and trust* [COM (2020) 65 final], p. 13.

¹⁸ BOUCHER, Philip; *Artificial intelligence: How does it work, why does it matter, and what can we do about it?*, European Parliamentary Research Service, Scientific Foresight Unit (STOA), 2020, Brussels.

¹⁹ Ya en 2020, HO, Daniel E., BLACK, Emily, AGRAWALA, Maneesh, y LI, Fei-Fei, profesores de la Facultad de Derecho de Stanford, publicaron «How regulators can get facial recognition technology right?», en la *Brookings Institution*. En el artículo instaban a los responsables políticos a abordar las cuestiones de intimidad y sesgo racial relacionadas con el reconocimiento facial. <https://www.brookings.edu/articles/how-regulators-can-get-facial-recognition-technology-right/>.

²⁰ SINGH, Aarav; «The Manifestation and Implications of Bias in Artificial Intelligence on Global Society», *International Journal For Multidisciplinary Research*, vol. 6/5, 2024. <https://doi.org/10.36948/ijfmr.2024.v06i05.28477>.

nerables, como mujeres²¹, minorías étnicas, personas con discapacidad o comunidades de bajos ingresos. Si los algoritmos se entrenan con datos procedentes principalmente de grupos favorecidos, los resultados también los beneficiarán a ellos. Al mismo tiempo, pueden excluir o perjudicar a los colectivos más vulnerables. Así, en lugar de corregir desigualdades, las refuerzan.

b) Sesgo de etiquetado: aparece cuando los datos con los que se entrena el algoritmo están mal etiquetados o solo lo están de forma parcial. El modelo aprende entonces patrones erróneos que pueden perjudicar especialmente a los grupos más vulnerables. Si los datos históricos ya reflejan prejuicios o discriminación, el algoritmo no solo los reproduce sino que los puede amplificar, perpetuando injusticias que ya existían.

c) Sesgo de representación: surge cuando falta diversidad en los datos o en las variables incluidas en el modelo. Si ciertos grupos (como las mujeres o las personas con discapacidad) no están bien representados, el algoritmo no podrá reconocer adecuadamente sus necesidades. El resultado puede ser una toma de decisiones que los excluya o los interprete de forma errónea.

d) Sesgo de medición: surge cuando las variables que se utilizan para entrenar el modelo no se miden de manera precisa, lo que afecta sobre todo a los grupos vulnerables. Si los métodos empleados para recopilar los datos no toman en cuenta la diversidad de los individuos o de los contextos, las decisiones algorítmicas se ven distorsionadas y acaban desfavoreciendo a esos grupos, perpetuando, una vez más, las desigualdades²².

e) Sesgo algorítmico de diseño: aparece cuando se priorizan algunas variables sin tener en cuenta contextos de vulnerabilidad. Por ejemplo, si un algoritmo de contratación valora sobre todo los años de experiencia, puede excluir a personas que no han tenido las mismas oportunidades laborales o educativas. Aunque estas personas cuenten con habilidades relevantes, el sistema no las detecta. El resultado es una exclusión indirecta.

f) Sesgo de retroalimentación: en ocasiones, los algoritmos producen resultados que ya son injustos. Si esos mismos resultados se usan para entrenar nuevos modelos, el sesgo se consolida. Es lo que se llama un ciclo de retroalimentación. En lugar de corregir desigualdades previas, el sistema las intensifica. La IA termina reforzando las mismas barreras que debería ayudar a superar.

²¹ La referencia a las mujeres dentro de los grupos en situación de vulnerabilidad no implica considerarlas vulnerables por naturaleza, sino reconocer que, en numerosos contextos jurídicos, sociales y tecnológicos, afrontan condiciones de vulnerabilidad estructural o contextual derivadas de desigualdades históricas y relaciones asimétricas de poder.

²² KUNDI, Bushra, EL MORR, Christo, GORMAN, Rachel y DUA, Ena, *Artificial Intelligence and Bias: A Scoping Review*, Chapman and Hall/CRC eBooks, 2022, pp. 199-215.

Por tanto, si se desea garantizar una IA inclusiva, es fundamental que los sistemas se diseñen de manera que promuevan la igualdad y consideren las necesidades específicas de los grupos vulnerables, contribuyendo así a la creación de una sociedad más justa. De ahí que la identificación y corrección de estos sesgos sea crucial para evitar que los algoritmos perpetúen o aumenten las discriminaciones sociales.

En el caso concreto de los sesgos de género, investigaciones a nivel internacional²³ demuestran que los sesgos de género que persisten en los conjuntos de datos, algoritmos y dispositivos de capacitación de la IA, tienen el potencial de propagar y reforzar estereotipos perjudiciales. Podrían, por tanto, llegar a estigmatizar aún más, por ejemplo, a las mujeres, con el peligro de quedar relegadas en varios ámbitos de la vida económica y profesional, entre otros. Así, la IA podría frenar los avances logrados en igualdad de género²⁴.

2.2.3 Casos de sesgos de género en aplicaciones de IA en el ámbito laboral

A pesar de los esfuerzos por reducir los prejuicios, se ha demostrado que la presencia de sesgos de género en el entrenamiento de diversos algoritmos de inteligencia artificial sigue existiendo²⁵. Y es que estos sesgos reflejan las estructuras culturales y sociales preexistentes. Un ejemplo especialmente ilustrativo es el uso de algoritmos en los procesos de selección de personal. A primera vista, pueden parecer herramientas neutras y objetivas, pero lo cierto es que también en este ámbito pueden generarse formas de discriminación que pasan desapercibidas. Algunas plataformas utilizan algoritmos para filtrar currículums y seleccionar candidaturas, pero si estos algoritmos se entrenan con datos sesgados o estandarizados, tienden a reproducir patrones que favorecen ciertos perfiles. Esto hace que se perpetúen desigualdades y se dificulte la diversidad en el lugar de trabajo. Asimismo, algunos estudios²⁶ han revelado que los algoritmos utilizados en plataformas de publicidad *on line* pueden mostrar

²³ UNESCO; *Preliminary study on the Ethics of Artificial Intelligence*, 2019, <https://unesdoc.unesco.org/ark:/48223/pf0000367823.locale=es>.

²⁴ RAMÍREZ AUTRÁN, Rodrigo; «Sesgos y discriminaciones sociales de los algoritmos en inteligencia artificial: una revisión documental», *Entretextos*, vol. 15/39, 2023, p. 7.

²⁵ MENDONÇA DE LIMA, Rosileine, R, PISKER, BARBARA AND CORRÊA, Victor Silva; *op. cit.*

²⁶ HE, Joyce, KANG, Sonia; «Identities between the lines: Re-aligning gender and professional identities by altering job advertisement language», *Academy of Management Proceedings*, vol. 2/1, 2022, p. 10415; WILSON, Kyra y CALISKAN, Aylin; «Gender, Race, and Intersectional Bias in Resume Screening via Language Model Retrieval», *Proceedings of the Seventh AAAI/ACM Conference on AI, Ethics, and Society (AIES2024)*, 2024, pp. 1578-1590; CHEN, Z. «Ethics and discrimination in artificial intelligence-enabled recruitment practices», *Humanities Social Sciences Communications*, vol. 10/567, 2023, pp. 1-12.

anuncios de empleo solo a un público específico, excluyendo a otros grupos de oportunidades laborales.

A continuación se presentan tres ejemplos concretos en el ámbito laboral que evidencian cómo determinados sesgos de género incorporados en sistemas algorítmicos han generado consecuencias negativas para las mujeres. Estos casos permiten mostrar cómo la falta de una perspectiva de justicia en sentido clásico en el diseño y aplicación de estas tecnologías puede perpetuar desigualdades ya existentes en lugar de contribuir a su superación²⁷.

A) Selección de personal

En 2014, Amazon desarrolló un sistema algorítmico con el fin de optimizar la asignación de recursos, reducir tiempos y mano de obra. La pretensión era crear un proceso de contratación considerado imparcial. Confiaban en desarrollar una IA capaz de analizar rápidamente la web y detectar a las candidaturas más adecuadas. El grupo desarrolló 500 modelos enfocados en funciones y ubicaciones laborales específicas, enseñándoles a identificar unos 50.000 términos comunes en los currículos de candidatos/as previos. Estos postulantes previos habían sido en su mayoría hombres, debido al dominio masculino en la industria tecnológica. Este sesgo en los datos condujo a decisiones distorsionadas por parte del algoritmo. Una de ellas fue que el sistema aprendió a preferir a candidatos masculinos. La compañía se vio obligada a modificarlo debido a que el algoritmo resultó sexista. No obstante, a pesar de las modificaciones, no se logró evitar el sesgo sexista²⁸ y Amazon finalmente abandonó el proyecto²⁹.

B) Anuncios de empleos

En 2021, una auditoría³⁰ realizada por investigadores independientes de la Universidad del Sur de California (USC) sobre anuncios de empleo

²⁷ *“Iustitia est constans et perpetua voluntas ius suum cuique tribuendi.* Cfr. Digesto, libro 1, título 1, fragmento 10, principio.

²⁸ DASTIN, Jeffrey «Insight – Amazon scraps secret AI recruiting tool that showed bias against women», Reuters, October 11, 2018, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/> WINICK, Erin; «Amazon ditched AI recruitment software because it was biased against women», *MIT Technology Review*, vol. 10, 2018, <https://www.technologyreview.com/2018/10/10/139858/amazon-ditched-ai-recruitment-software-because-it-was-biased-against-women/>.

²⁹ DASTIN, Jeffrey, «Amazon abandona un proyecto de IA para la contratación por su sesgo sexista», Reuters, October 14, 2018, (trad. Tomás Cobos), <https://www.reuters.com/article/world/amazon-abandona-un-proyecto-de-ia-para-la-contratacin-por-su-sesgo-sexista-idUSKCN1MO0M4/>.

³⁰ IMANA, Basileal., KOROLOVA, Aleksandra y HEIDEMANN, John; «Auditing for Discrimination in Algorithms Delivering Job Ads», *Actas de The Web Conference 2021 (WWW '21)*, del 19 al 23 de abril de 2021, Liubliana, Eslovenia, 2021, <https://ant.isi.edu/datasets/addelivery/>.

reveló que el sistema de anuncios de Facebook mostraba diferentes anuncios a mujeres y hombres, aunque los trabajos requiriesen la misma cualificación. Los resultados del trabajo mostraban que los algoritmos de distribución de anuncios de Facebook mantenían patrones discriminatorios de género en la contratación, a pesar de las promesas de la empresa de abordar este problema³¹. El estudio reveló que los anuncios de trabajos con características similares se mostraban a diferentes grupos demográficos según su sexo, reproduciendo patrones históricos de empleo, como en los casos de trabajos de repartidor, ventas de automóviles o de joyas³². A pesar de las demandas y las auditorías previas que evidenciaron sesgos, Facebook no había introducido cambios sustanciales en su sistema de anuncios. No obstante, el mismo trabajo concluía que no había localizado esos sesgos en LinkedIn.

C) Cartas de recomendación

En los últimos años, los modelos de lenguaje de gran tamaño (*Large Language Models*, LLM) han comenzado a utilizarse en distintos entornos profesionales y académicos. Su capacidad para generar textos largos y estructurados ha abierto nuevas posibilidades de uso. Una de ellas es la redacción automática de cartas de recomendación.

En principio, esta opción puede resultar útil porque podría ahorrar tiempo y ofrecer borradores aptos para quienes deben redactar estos documentos para entornos laborales. El ahorro de tiempo y el acceso a una redacción formal pueden parecer ventajas evidentes. Sin embargo, conviene analizar esta práctica con cuidado. No se trata solo de redactar bien, sino de valorar si el perfil de una persona encaja en el ámbito en el que aplica. Y ahí es donde surgen los primeros interrogantes porque esta aparente neutralidad técnica no está exenta de riesgos.

En 2023, investigadores de la Universidad de California llevaron a cabo un estudio³³ en el que solicitaron a dos modelos de IA: ChatGPT (de OpenAI) y Alpaca (un modelo de lenguaje entrenado por la Universidad de Stanford) que generasen cartas de recomendación para perfiles ficticios. Los resultados revelaron un patrón preocupante: ambos modelos tendían a reproducir estereotipos de género en la selección de adjetivos y sustantivos. Por ejemplo, ChatGPT describía

³¹ ProPublica sacó a la luz el problema en 2016 (ANGWIN Julia y PARRIS, Terry Jr.; «Facebook Lets Advertisers Exclude Users by Race», *Propublica*, 28 de octubre 2016 (<https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race>)).

³² Los investigadores encontraron el mismo patrón con anuncios para trabajos: ingenieros de software para Nvidia (predominantemente hombres) y Netflix (predominantemente mujeres), y vendedores de automóviles (predominantemente hombres) y joyas (predominantemente mujeres).

³³ WAN, Yixin, PU, George, SUN, Jiao, GARIMELLA, Aparna, CHANG, Kai-Wei y PENG, Nanyun, «Kelly is a Warm Person, Joseph is a Role Model: Gender Biases in LLM-Generated Reference Letters», *ArXiv*, Cornell University, 2023, <https://arxiv.org/pdf/2310.09219v5>.

con mayor frecuencia a los hombres como «expertos» o «íntegros», mientras que a las mujeres se las asociaba con términos como «belleza» o «delicia». Algo similar ocurría con Alpaca: los hombres eran «pensadores» y «oyentes», mientras que las mujeres aparecían como «gráciles» o «emocionales».

Estos sesgos no son anecdóticos. Reflejan cómo la IA, cuando no es adecuadamente entrenada, puede acabar replicando —e, incluso, reforzando— estereotipos profundamente arraigados. De ahí la necesidad de abordar la cuestión, especialmente cuando estas tecnologías empiezan a intervenir en procesos que pueden condicionar trayectorias profesionales.

3. LOS PLANES DE IGUALDAD

En España, desde marzo de 2022, todas las empresas con más de cincuenta personas en plantilla están obligadas a contar con un Plan de Igualdad. Así lo establece la Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva entre mujeres y hombres.

Estos planes son más que una obligación formal. Se conciben como instrumentos prácticos para avanzar hacia una igualdad real en el entorno laboral. Se elaboran a partir de un diagnóstico previo que identifica las desigualdades existentes en la empresa. A partir de ese diagnóstico, el plan debe incluir unos objetivos concretos, unas medidas específicas para reducir las desigualdades y un sistema de seguimiento que permita evaluar su cumplimiento. Las áreas que se abordan son la contratación, la formación, la promoción interna, la conciliación y la brecha salarial.

Más allá de su valor jurídico, los Planes de Igualdad representan una oportunidad real para promover entornos laborales libres de discriminación. No se trata solo de cumplir con una obligación legal, sino de transformar las dinámicas internas de las organizaciones. La igualdad deja de ser un principio abstracto para convertirse en una práctica concreta, medible y orientada al cambio.

Ya se ha señalado que su objetivo es garantizar la igualdad de trato y oportunidades entre mujeres y hombres. Por eso, y teniendo en cuenta los sesgos de género que pueden presentar algunas herramientas de inteligencia artificial en el ámbito laboral, estos planes pueden convertirse en una herramienta clave para regular y supervisar el uso de la IA. En particular, pueden actuar de forma preventiva, evitando que los sistemas automatizados perpetúen desigualdades ya existentes.

A continuación se proponen dos líneas de actuación complementarias: 1) La primera se centra en los propios sistemas: es necesario establecer directrices claras para la creación y evaluación de algoritmos, de modo que se garantice su vinculación con los principios de igual-

dad entre mujeres y hombres; 2) La segunda línea hace referencia a la formación: es imprescindible dotar al personal de conocimientos que permitan entender, supervisar y aplicar la inteligencia artificial de forma justa.

3.1 Medidas de transparencia e igualdad en la creación y evaluación de algoritmos

Incorporar la transparencia y la supervisión de los algoritmos en los Planes de Igualdad es un paso imprescindible. Solo así podrá evitarse que la inteligencia artificial reproduzca –o incluso aumente– los sesgos de género ya existentes. Como se ha indicado, los algoritmos se entrenan con datos que, en ocasiones, reflejan desigualdades históricas. Si no se gestionan con cuidado, pueden acabar reforzando estereotipos y formas de discriminación que se pretendían superar. Los Planes de Igualdad, al ser su objetivo garantizar la igualdad entre mujeres y hombres, ofrecen un marco adecuado para incorporar medidas que vigilen el uso correcto de los algoritmos en las empresas. Estas herramientas no deben favorecer ni discriminar a nadie por razón de sexo. La transparencia resulta clave: permite detectar sesgos y corregirlos antes de que afecten a procesos como la selección o la promoción profesional. A su vez, contribuye a generar confianza, tanto en las personas candidatas como en el personal en plantilla. Incluir medidas de supervisión algorítmica refuerza ese compromiso. Es una forma concreta de asegurar que la tecnología contribuya a la inclusión.

Un ejemplo concreto es el de la valoración de las competencias socioemocionales (SEL). Estas competencias abarcan conocimientos, habilidades y actitudes que permiten reconocer y gestionar las propias emociones, comprender a los demás y establecer relaciones prosociales³⁴. Suelen agruparse en cinco dimensiones: autoconciencia, automotivación y autogestión, conciencia social, conducta prosocial y toma de decisiones responsable³⁵.

Tradicionalmente, este tipo de habilidades –las denominadas *soft skills*– se han asociado a las mujeres. Por eso, una medida clave para reducir la discriminación de género en los algoritmos consiste en garantizar, dentro del Plan de Igualdad, que estas competencias no sean infravaloradas frente a las habilidades técnicas, más vinculadas culturalmente a los hombres.

³⁴ Esta perspectiva ha sido respaldada por la *Collaborative for Academic, Social and Emotional Learning* (CASEL). Para más información, *vid.* DURLAK, Joseph A., DOMITROVICH, Celene E. y MAHONEY, Joseph L.; *Handbook of Social and Emotional Learning*, 2.º Ed., Guilfor Press, 2024.

³⁵ ŻYCH, Izabela, BELTRÁN-CATALÁN, María, ORTEGA-RUIZ, Rosario y LLORENT, Vicente J.; «Social and emotional competencies in adolescents involved in different bullying and cyberbullying roles», *Revista de Psicodidáctica*, vol. 23/2, 2018, pp. 86-93.

Es importante recordar que los algoritmos de aprendizaje automático replican los sesgos presentes en los datos con los que se entrenan. Si el modelo aprende que cualidades como la empatía, la colaboración o la comunicación son más frecuentes en mujeres, es posible que las valore como menos «profesionales». En cambio, dará mayor peso a habilidades tradicionalmente masculinas, como la toma de decisiones basada en datos o la dirección estratégica. Supervisar estos sesgos es esencial para evitar que se perpetúe una visión desequilibrada del talento³⁶.

Desde una perspectiva técnica, esto implica la necesidad de recalibrar los modelos para que las *soft skills* sean igualmente reconocidas y valoradas, independientemente del sexo. Al no hacerlo, los algoritmos pueden descalificar injustamente a candidatas y candidatos cualificados, lo que impactaría negativamente en la diversidad y efectividad organizacional. Por lo tanto, educar al algoritmo para reconocer las *soft skills* es crucial para una toma de decisiones justa y para contribuir a lograr buenos ambientes laborales.

El papel de las competencias socioemocionales en el ámbito de la inteligencia artificial no debe infravalorarse. Comprender las implicaciones sociales de cada decisión algorítmica permite ampliar el foco más allá de lo técnico, considerando también el impacto que estos sistemas pueden tener en las personas y en las comunidades. Esta mirada favorece una IA más justa.

3.2 Medidas de formación en igualdad

Las medidas de formación en igualdad cumplen una función clave. No solo sensibilizan sobre la falta de igualdad real entre mujeres y hombres, sino que ayudan a reconocer que el sexismo sigue presente en muchas dinámicas sociales y profesionales. Esta formación busca fomentar una actitud activa frente a cualquier forma de discriminación.

A menudo, estas acciones formativas se centran en detectar estereotipos de género vinculados al entorno laboral. También abordan la corresponsabilidad como un factor necesario, tanto para el bienestar de las personas empleadas como para la salud y sostenibilidad de la empresa u organización. Por tanto, las medidas de formación en igualdad no son accesorias: son una herramienta esencial para avanzar hacia una igualdad efectiva y duradera en los entornos laborales³⁷. Estas medidas pueden facilitar que quienes supervisan la aplicación de los Planes de Igualdad detecten posibles sesgos, incluidos los que puedan derivarse del uso de herramientas de IA. Contar con esta capacidad de análisis es esencial para corregir desigualdades y avanzar

³⁶ Recuérdese el ejemplo de ChatGpt y las cartas de recomendación.

³⁷ DÍAZ DE TERÁN, María Cruz. (coord.); *Formación en igualdad: un enfoque desde La Ley 3/2007, de 22 de marzo, de igualdad efectiva de mujeres y hombres*, Navarra, Thomson-Reuters, 2022, p. 15.

hacia un entorno laboral que garantice la igualdad de oportunidades para mujeres y hombres.

Junto con ello, la formación en igualdad del personal técnico que desarrolla la IA es también fundamental. Esta formación debe combinar la teoría con la práctica, estar orientada a identificar puntos críticos en el diseño de sistemas, para contribuir así a reforzar la transparencia, anticipar los efectos negativos de los sesgos y mitigar sus consecuencias.

4. ÉTICA Y RESPONSABILIDAD EN EL USO DE HERRAMIENTA DE IA

Sin duda, el desarrollo acelerado de la inteligencia artificial plantea importantes interrogantes ético-jurídicos. Como ha subrayado Adela Cortina, uno de los primeros temas que deben ponerse sobre la mesa es «la diferencia abismal que existe entre “hacer” uso de sistemas inteligentes (sean máquinas, algoritmos, robots) a la hora de tomar decisiones y “delegar” en esos sistemas inteligentes decisiones significativas para la vida de las personas y de la naturaleza»³⁸. La cuestión, por tanto, no es si avanzar en esta dirección, sino cómo hacerlo.

Diversos organismos internacionales están asumiendo la responsabilidad de establecer marcos éticos para guiar este proceso. En noviembre de 2021, los 193 Estados miembros de la UNESCO adoptaron la primera Recomendación universal sobre la ética de la inteligencia artificial³⁹. En el ámbito europeo, destacan las Directrices éticas para una IA fiable elaboradas por la Comisión Europea (2019)⁴⁰ y la Carta ética europea sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno (2018)⁴¹. Todos estos instrumentos subrayan la importancia de una IA confiable, entendida como una tecnología que genere confianza y contribuya a una «competitividad responsable». La confianza, tanto técnica como ética, se presenta como un requisito previo para cualquier desarrollo significativo.

4.1 Principios ético-jurídicos fundamentales

En el contexto normativo europeo, más allá de la reciente Ley de Inteligencia artificial (Reglamento (UE) 2024/1689 por el que se esta-

³⁸ Cfr. CORTINA, Adela; «Ética de la inteligencia artificial», *Anales de la Real Academia de Ciencias Morales y Políticas*, vol. 96, 2019, p. 381, https://www.boe.es/biblioteca_juridica/anuarios_derecho/abrir_pdf.php?id=ANU-M-2019-10037900394.

³⁹ https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa.

⁴⁰ <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>.

⁴¹ <https://protecciondata.es/wp-content/uploads/2021/12/Carta-Etica-Europea-sobre-el-uso-de-la-Inteligencia-Artificial-en-los-sistemas-judiciales-y-su-entorno.pdf>.

blecen normas armonizadas sobre inteligencia artificial)⁴², la Unión Europea ha ido construyendo un marco normativo que busca orientar el desarrollo de estas nuevas tecnologías desde una perspectiva ético-jurídica. El punto de partida está claro: no se trata solo de garantizar la competitividad digital en el contexto global, sino de asegurar que esta transformación tecnológica se mantenga fiel a los valores que definen el proyecto europeo: la dignidad humana, la justicia y el respeto a los derechos fundamentales⁴³.

Entre los instrumentos clave destacan las ya mencionadas Directrices éticas para una IA fiable, elaboradas por el Grupo de expertos de alto nivel sobre inteligencia artificial (HLEG AI, 2019)⁴⁴. Este documento establece un marco ético-jurídico que apuesta por sistemas tecnológicamente robustos, éticamente sólidos y socialmente responsables⁴⁵. A su vez, el Libro blanco sobre inteligencia artificial. Un enfoque europeo orientado a la excelencia y la confianza (2020)⁴⁶, marca una hoja de ruta en la que se suman la promoción de la innovación, la gestión de riesgos y la elaboración de un modelo regulatorio propio⁴⁷.

De la trayectoria de los documentos europeos, puede constarse que la estrategia europea de inteligencia artificial se asienta en el principio de que la IA debe orientarse al servicio de la sociedad en su conjunto, garantizando que sus aplicaciones contribuyan efectivamente al bienestar colectivo y no aumenten las desigualdades. En este sentido, el Reglamento general de protección de datos (GDPR)⁴⁸, aunque anterior al auge de la IA generativa, sigue desempeñando un papel crucial para asegurar que los sistemas inteligentes respeten los derechos fundamentales, en particular la intimidad y la autodeterminación informativa⁴⁹.

⁴² <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32024R1689>.

⁴³ En el punto (8) del Preámbulo se indica: «En consecuencia, se necesita un marco jurídico de la Unión que establezca unas normas armonizadas en materia de IA para impulsar el desarrollo, la utilización y la adopción en el mercado interior de la IA y que, al mismo tiempo, ofrezca un nivel elevado de protección de los intereses públicos, como la salud y la seguridad y la protección de los derechos fundamentales, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, reconocidos y protegidos por el Derecho de la Unión».

⁴⁴ <https://digital-strategy.ec.europa.eu/es/library/ethics-guidelines-trustworthy-ai>.

⁴⁵ Sobre estas directrices, *vid.*, COLCELLI, Valentina y BURZAGLI, Laura; «Elementos para una cultura europea de desarrollo de herramientas de inteligencia artificial: el Libro blanco sobre la inteligencia artificial y las Directrices éticas para una IA fiable», *Revista Justicia y Derecho*, vol. 4/2, 2021, pp. 1-12.

⁴⁶ <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065>.

⁴⁷ LARSSON, Stefan; «On the Governance of Artificial Intelligence through Ethics Guidelines», *Asian Journal of Law and Society*, vol. 7/3, 2020, pp. 437-451.

⁴⁸ <https://www.boe.es/doue/2016/119/L00001-00088.pdf>.

⁴⁹ HERNÁNDEZ, Juan Carlos, «Gobernanza de la inteligencia artificial en la Unión Europea. La construcción de un marco ético-jurídico aún inacabado», *Revista General de Derecho Administrativo*, vol. 56, 2021; EKE, Damian y STAHL, Bernd; «Ethics in the Governance of Data and Digital Technology: an Analysis of European Data Regulations and Policies», *Digital Society*, vol. 3/11, 2024, pp. 1-23.

En junio de 2025, el Consejo de la Unión Europea reforzó esta línea con la adopción de las Conclusiones sobre el fomento de la igualdad de género en la era digital impulsada por la inteligencia artificial: sexta revisión horizontal de la aplicación de la Plataforma de Acción de Pekín⁵⁰. Este documento –de carácter vinculante en términos de orientación política– constituye un hito en la incorporación explícita de la perspectiva de género al debate sobre la gobernanza de la IA. En su apartado 3, dedicado al sesgo de género en los sistemas de inteligencia artificial y a los mecanismos de supervisión, el Consejo reconoce que la IA, si no se somete a un control adecuado, puede reproducir y amplificar discriminaciones estructurales existentes. Por ello, insta a los Estados miembros y a las instituciones europeas a garantizar el uso de datos representativos, de calidad y desglosados por sexo y otras variables interseccionales en el diseño, entrenamiento y evaluación de los sistemas automatizados. Asimismo, subraya la necesidad de introducir supervisión humana efectiva (*human oversight*) en los procesos algorítmicos, tanto para detectar posibles sesgos como para corregir los resultados que puedan afectar a la igualdad de trato.

Estas Conclusiones también enfatizan que el potencial de la inteligencia artificial para contribuir a la igualdad dependerá de la existencia de mecanismos transparentes, auditables y conformes con el Derecho antidiscriminatorio de la Unión, de manera coherente con los principios recogidos en la Ley de IA (2024/1689) y en la Estrategia de Igualdad de Género de la UE (2020-2025). El texto del Consejo invita, además, a promover investigaciones sobre los impactos diferenciales de la automatización y a integrar indicadores de género en la evaluación de impacto de las políticas digitales.

En su conjunto, estos documentos configuran una estructura ético-jurídica que pretende guiar la inteligencia artificial desde una lógica de responsabilidad, anticipación y justicia. No se trata únicamente de evitar los riesgos, sino de construir un modelo de desarrollo tecnológico compatible con la democracia, la libertad y los derechos fundamentales.

4.2 Filosofía del Derecho, ética y sesgos en IA

Ya se ha indicado que los sistemas de IA aprenden a partir de los datos que reciben, por tanto, si los datos contienen sesgos, estos pueden no solo replicarse sino incluso, aumentarse. En contextos especialmente sensibles como el sexo, la etnia o la clase social, este fenómeno puede dar lugar a formas sutiles pero persistentes de discriminación.

La gestión ética de estos sesgos implica prestar atención no solo a los datos, sino también a los modelos y a las personas que intervienen en su diseño. El impacto de los sesgos en decisiones que afectan direc-

⁵⁰ ST 9984/25, Consejo EPSCO, 19 de junio de 2025.

tamente a la vida de las personas exige respuestas. Aunque no todo sesgo implica una discriminación jurídica, sí puede generar situaciones de injusticia, especialmente cuando se da un trato diferenciado sin una justificación objetiva⁵¹. Diseñar algoritmos inclusivos exige desarrollar estrategias de corrección de sesgos desde un enfoque interdisciplinar y colaborativo. Junto con ello, las empresas tecnológicas, las instituciones educativas, la Administración y el conjunto del tejido social tienen un papel que desempeñar. La incorporación de políticas de igualdad en la formación técnica, así como en el diseño y validación de modelos de IA, resulta imprescindible.

La Filosofía del Derecho debe explorar con profundidad las relaciones entre ética, inteligencia artificial y Derecho. Esta disciplina tiene mucho que aportar ante los retos que plantea la IA. No solo debe reflexionar sobre sus implicaciones ético-jurídicas, sino también detectar mecanismos que favorezcan la justicia. En este marco, los Planes de Igualdad ofrecen una oportunidad concreta de intervención. Si los algoritmos reproducen desigualdades o estereotipos de género, los Planes permiten supervisar y corregir esos sesgos. Integrar esta perspectiva en la investigación filosófico-jurídica permite actualizar conceptos como justicia o igualdad en una sociedad digital. Además, refuerza el valor práctico de la Filosofía del Derecho y su compromiso con los derechos humanos en contextos tecnológicos⁵².

5. CONCLUSIONES

El avance de la inteligencia artificial (IA) es imparable. Su impacto en múltiples ámbitos de la vida cotidiana, del trabajo y de la toma de decisiones es ya una realidad. Precisamente por eso resulta urgente establecer un marco ético-jurídico sólido que oriente su desarrollo y aplicación. No basta con regular los efectos de la IA *a posteriori*. Entre las medidas principales destacan: anticiparse a sus posibles consecuencias, integrar la perspectiva de género desde el diseño y asegurar que los derechos fundamentales de todas las personas (especialmente de quienes históricamente han sido grupos vulnerables) sean respetados desde el inicio.

En este sentido, los Planes de Igualdad pueden convertirse en una herramienta eficaz para incorporar esta mirada con carácter transversal, promoviendo entornos tecnológicos más justos e inclusivos. De ahí que para promover una integración responsable de la inteligencia

⁵¹ MARTÍN LÓPEZ, Jorge; «Inteligencia artificial, sesgos...», p. 65.

⁵² Sobre los posibles riesgos en el ámbito de los derechos humanos, *vid.*, AÑÓN, María José; «Desigualdades algorítmicas: conductas de alto riesgo para los derechos humanos», *Derechos y libertades: Revista de Filosofía del Derecho y derechos humanos*, vol. 47, 2022, pp. 17-40.

artificial, comprometida con la igualdad entre mujeres y hombres sea imprescindible establecer procesos de evaluación sistemáticos que permitan revisar tanto los datos de entrada como los resultados generados por los sistemas de IA. Esta supervisión constante facilitará la detección temprana de sesgos de género, impidiendo su consolidación y posterior reproducción.

Junto con ello, para evitar que la IA perpetúe desigualdades estructurales, los algoritmos deben nutrirse de datos diversos, desglosados por sexo y sensibles a la pluralidad de realidades sociales. La representatividad en los datos no es un detalle técnico: es una exigencia de justicia.

Además, la creación de procedimientos claros, que combinen herramientas automatizadas con la intervención de personas expertas en igualdad, permitirá responder de forma más eficaz ante sesgos persistentes. Esta tarea requiere también una asignación específica de recursos, tanto técnicos como humanos.

Otra clave es la participación activa de las personas usuarias y trabajadoras. Establecer canales accesibles para reportar posibles sesgos en los sistemas algorítmicos utilizados por las organizaciones permitirá avanzar hacia modelos más transparentes, abiertos al aprendizaje y a la mejora continua.

La integración de estos pasos en los Planes de Igualdad constituye una oportunidad concreta para transformar los entornos laborales y garantizar que la inteligencia artificial no se convierta en una nueva fuente de discriminación, sino en una herramienta aliada de una sociedad más ética, más humana y más justa.

