

Fuentes de sesgos en inteligencia artificial: principios básicos para evitarlos en el área de RRHH/Inteligencia artificial y decisión

Carlos Enrique Ezeiza Pohl¹

Resumen

Entre los principales retos que presenta la gobernanza de la IA en la actualidad se encuentran las fuentes de sesgos que afectan a los distintos modelos de IA, tanto en el caso del razonamiento automático como en el del Machine Learning. Identificar adecuadamente los sesgos y propiciar buenas prácticas para evitarlos constituye uno de los desafíos críticos que afectan al desarrollo y uso de aplicaciones en IA. En este trabajo, de carácter teórico, se presenta una revisión sistemática de literatura sobre las fuentes de sesgos en IA y su aplicación en la toma de decisiones basadas en IA en el área de búsqueda y selección de personal. Se relevan e identifican los principales sesgos detectados (algorítmicos, cognitivos, confirmación, exclusión, medición, homogeneidad del exogrupo, prejuicio, recuerdo, muestra/selección, estereotipos), y los elementos para tener en cuenta a fin de evitarlos (modelos de aprendizaje supervisados y no supervisados; datos de procedencia diversa, completos y equilibrados; composición diversificada del staff que diseña y mantiene la aplicación de IA en cuanto a género, edad, experiencia, etnia y multiculturalidad; monitoreo constante e infraestructura de TI). Se presentan los resultados parciales de una revisión sistemática de literatura de casos de sesgos detectados en aplicaciones de IA utilizadas para la búsqueda y selección de personal y los aspectos para tener en cuenta con el propósito de reducir su impacto y consecuencias para las organizaciones y la sociedad en su conjunto.

Palabras claves: sesgos - inteligencia artificial - recursos humanos - selección de personal

Introducción

Referirnos a que la Inteligencia Artificial (IA) domina una gran cantidad de espacios, experiencias y realidades de la experiencia humana en la actualidad es casi caer en un lugar común, y más bien habría que excluir los pocos casos o ámbitos en los que aún no ha producido

un impacto significativo, algo similar a lo que sucede hoy en día con la ubicuidad que presenta el uso de dispositivos informáticos y sus aplicaciones, ya preanunciada por la Revolución Industrial 4.0 que desde la primera década del presente siglo se manifiesta a través de la introducción en la actividad productiva de sistemas

¹ Docente-Investigador. Director Ejecutivo de la Revista RInCE Revista de Investigaciones del Departamento de Ciencias Económicas de la Universidad Nacional de La Matanza (UNLaM)

de inteligencia artificial y procesos de automatización e interconexión del trabajo, y en la cual es clave la gestión y explotación de datos, la interacción entre el mundo físico y el digital, el aprendizaje artificial autónomo y la introducción de dispositivos híbridos de *hardware* y *software* capaces de aprender y de desempeñar tareas cada vez más complejas (Choi, 2021).

En este escenario actual de ubicuidad de la gestión y explotación de datos, y de interacción entre el mundo físico y el digital, especialmente a través de la telefonía móvil —denominación de por sí ya obsoleta porque el teléfono celular es una computadora portátil que cada vez se emplea menos para hablar en tiempo real y sincrónica, y más para la comunicación asincrónica y de interacción a través de redes sociales y la gestión integral de información— cobra sentido lo que Kay y Goldberg preanunciaron en 1977 a través de su teoría de la computadora como “metamedio”, es decir, como principal tecnología de comunicación, información y convergencia, categoría a la cual se puede aplicar al *smartphone* en nuestros días (Márquez, 2017). En este sentido, tampoco la IA es una tecnología que sea exclusiva de esta época, sus desarrollos iniciales bien pueden rastrearse a través de los trabajos pioneros de Alan Turing en 1935, y luego en los laboratorios en Bletchley durante la Segunda Guerra Mundial (Valor Yébenes, 2024), pero claramente desde la Revolución 4.0 se ha demostrado que su alcance es cada

vez mayor en amplitud y tareas en la que interviene, y el mercado laboral es una de ellas, de la cual nos ocuparemos en este trabajo.

En toda organización se encuentra permanentemente la necesidad de atender los requerimientos de incorporación, promoción, retención e inclusive de desvinculación de su dotación de personal, actividad de la cual es responsable el sector de Recursos Humanos (RR. HH.), pero cuya transcendencia supera los límites departamentales de dicho sector e involucra estratégicamente a toda la estructura y en especial a su nivel directivo (Chiavenato, 2011), y actualmente se ha hecho manifiesta la tendencia en el uso de aplicaciones de IA en los procesos de búsqueda y reclutamiento de personal. En este trabajo se propone analizar —a través de una revisión sistemática de literatura— las fuentes de sesgos que presenta el uso de IA en la toma de decisiones respecto de la búsqueda y selección de personal. En una primera parte se examinan los principales sesgos que presenta el uso la IA (algorítmicos, cognitivos, confirmación, exclusión, medición, homogeneidad del exogrupo, prejuicio, recuerdo, muestra/selección, estereotipos); seguido de una segunda parte en donde se analizan los factores por tener en cuenta para reducir su impacto, y en una tercera parte se presenta una revisión sistemática de literatura en casos en donde se realiza una búsqueda de personal utilizando IA y se manifiestan los sesgos antes men-

cionados. Finalmente, en la conclusión se extraen reflexiones sobre los riesgos y estrategias para aplicar en la búsqueda de personal al utilizar IA, a fin de minimizar el impacto de los sesgos en la toma de decisiones al seleccionar personal en una organización.

1. Sesgos presentes en el uso de la IA

En primer lugar, la definición de sesgo proviene principalmente del campo de la estadística —con la cual la IA presenta una gran integración y relación interdisciplinar— y se puede considerar que existe sesgo cuando en el curso de una recolección, procesamiento y análisis de datos se comete un error de forma sistemática, es decir, no aparece como un hecho aleatorio o por efecto del azar (Manterola y Otzen, 2015). En la literatura consultada se reconocen alrededor de diez categorías de sesgos que presenta el uso de IA, las cuales se reconocen como: algorítmicos, cognitivos, confirmación, exclusión, medición, homogeneidad del exogrupo, prejuicio, recuerdo, muestra/selección, estereotipos. A continuación, desarrollaremos cada uno de ellos a partir de la consulta al artículo "*What is AI bias?*" publicado por la empresa tecnológica estadounidense IBM, y que disponemos en la Tabla 1.

A continuación, se presentan los principios para tener en cuenta a fin de reducir el impacto de los sesgos mediante el uso de aplicaciones de IA.

2. Principios para tener en cuenta a fin de reducir el impacto de los sesgos en el uso de la IA

Con base en el artículo "*What is AI bias?*", las recomendaciones para reducir el impacto del sesgo en sus diferentes manifestaciones tal y como fueron presentados en la Tabla 1, se centran conceptos como gobernanza de la IA, en el sentido de promover un equilibrio entre los beneficios que la IA otorga a las empresas, los clientes, los empleados y la sociedad en su conjunto (Holdsworth, 2025). Por otra parte, Monasterio Astobiza (2021) da cuenta de que la gobernanza de la IA incluye generalmente métodos que buscan evaluar la imparcialidad, la equidad y la inclusión en la toma de decisiones que resultan del uso de aplicaciones de IA, en tanto, enfoques como la imparcialidad contrafactual se ocupan de identificar los sesgos que un modelo presenta en la toma de decisiones y garantizan resultados equitativos, incluso cuando se incluyen atributos sensibles como el género, la raza o la orientación sexual. Reducir el impacto de los sesgos requiere dar cuenta de la complejidad que presenta la IA, en donde un algoritmo puede ser considerado un sistema de caja negra con poca información sobre los datos que se han utilizado para crearlo, por cuanto las prácticas y tecnologías de transparencia de gobernanza que se utilicen permitirán garantizar que se empleen datos imparciales para construir el sistema y que los resultados sean justos (Ramírez Autrán, 2023).

Tabla 1. Sesgos presentes en el uso de IA

N.º	Denominación	Definición	Interviene
1	Algorítmico	Se presenta si el problema o la pregunta formulada no es totalmente correcta o específica, o si la retroalimentación al algoritmo de aprendizaje automático no ayuda a guiar la búsqueda de una solución.	Tanto en la formulación de la pregunta/problema como en la fuente de datos que provee al algoritmo.
2	Cognitivo	Nuestros valores, creencias, experiencias y modelos culturales condicionan nuestra manera de pensar y atribuir significado a la realidad de forma permanente, aún sin que nos demos cuenta.	Tanto al conjunto de datos de origen como al comportamiento del modelo de IA.
3	Confirmación	Relacionado con el sesgo anterior, aparece cuando la IA se basa redundantemente en creencias o tendencias preexistentes en los datos, al duplicar los sesgos existentes y sin poder identificar nuevos patrones o tendencias.	Tanto al conjunto de datos de origen como al comportamiento del modelo de IA.
4	Exclusión	Aparece cuando se omiten datos importantes de la fuente de datos que se utiliza, a menudo porque el desarrollador no ha logrado ver factores nuevos e importantes.	Intervención del desarrollador en el conjunto de datos.
5	Medición	Proviene de datos incompletos en la fuente o set de datos, debido a un descuido o falta de preparación, lo que resulta en que el conjunto de datos no incluya a toda la población que debería considerarse.	Omisión por sesgo del responsable de preparar el set de datos que provee al modelo de IA.
6	Homogeneidad del exogrupo	Existe una tendencia a comprender mejor a los miembros del grupo al que se pertenece y a pensar que son más diversos que los del exogrupo . Como resultado, los desarrolladores pueden crear algoritmos con menor capacidad para distinguir entre individuos que no forman parte del grupo mayoritario en los datos de entrenamiento.	Omisión por parte del desarrollador, lo que genera sesgo racial, clasificación errónea y respuestas incorrectas.
7	Prejuicio	Aparece cuando estereotipos y suposiciones sociales afectan el conjunto de datos del algoritmo, lo que inevitablemente genera resultados sesgados. Por ejemplo, categorizar a una profesión o actividad como masculina o femenina.	Tanto al conjunto de datos de origen como al comportamiento del modelo de IA.
8	Recuerdo	Asociado al sesgo cognitivo, aparece por prevalencia subjetiva en el etiquetado de datos, donde las etiquetas se aplican de manera inconsistente según las observaciones subjetivas.	El equipo responsable de supervisar el conjunto de datos de origen.
9	Muestra/selección	Aparece cuando los datos utilizados para entrenar el modelo de aprendizaje automático no son lo suficientemente amplios, representativos o completos como para entrenar el sistema adecuadamente.	El equipo responsable de supervisar el conjunto de datos de origen.
10	Estereotipos	Aparece cuando la IA refuerza estereotipos étnicos, etarios o de género.	Tanto al conjunto de datos de origen como al comportamiento del modelo de IA.

Fuente: elaboración propia con base en UN-COMTRADE. Miles de USD FOB.

En principio, y a modo preliminar, se presenta una lista de verificación de seis pasos del proceso que contribuye a reducir el impacto de sesgos en aplicaciones de IA, según se observa en la Tabla 2. A continuación, se revisa un caso en

el que se utilizó IA en la selección de personal, y se revisa según la literatura consultada los factores para tener en cuenta a fin de reducir el impacto de los sesgos de la IA al seleccionar personal.

Tabla 2. Seis pasos para reducir el impacto en el uso de IA

N.º	Paso	Descripción
1	Seleccionar el modelo de aprendizaje correcto	- Al utilizar un modelo supervisado, las partes interesadas seleccionan los datos de entrenamiento. Es fundamental que el equipo de partes interesadas sea diverso —no solo científicos de datos— y que haya recibido capacitación para prevenir sesgos inconscientes. - Los modelos no supervisados utilizan únicamente IA para identificar sesgos. Es necesario integrar herramientas de prevención de sesgos en la red neuronal para que aprenda a reconocer los sesgos.
2	Revisión de carga de datos en el modelo	Es necesario integrar herramientas de prevención de sesgos en la red neuronal para que aprenda a reconocer los sesgos.
3	Integración de equipo (<i>staff</i>)	Cuanto más diverso sea el equipo de IA (racial, económica, por nivel educativo, por género y por descripción del puesto), mayor será la probabilidad de que detecte sesgos. Los talentos y puntos de vista de un equipo de IA integral deben incluir innovadores empresariales de IA, creadores de IA, implementadores de IA y una representación de los consumidores de este proyecto de IA en particular.
4	Revisión constante del procesamiento de datos	Las empresas deben ser conscientes de los sesgos en cada paso del procesamiento de datos. El riesgo no solo reside en la selección de datos: ya sea durante el preprocesamiento, el procesamiento durante o el posprocesamiento, los sesgos pueden aparecer en cualquier momento y transmitirse a la IA.
5	Monitoreo constante de datos	El monitoreo y las pruebas constantes con datos reales de toda la organización pueden ayudar a detectar y corregir sesgos antes de que causen daños. Para evitar aún más sesgos, las organizaciones deben considerar evaluaciones realizadas por un equipo independiente dentro de la organización o por un tercero de confianza.
6	Datos recolectados de distinta infraestructura	Además de la influencia humana y de los datos, a veces la propia infraestructura puede generar sesgos. Por ejemplo, al utilizar datos recopilados por sensores mecánicos, el propio equipo podría generar sesgos si los sensores no funcionan correctamente. Este tipo de sesgo puede ser difícil de detectar y requiere inversión en infraestructuras digitales y tecnológicas de vanguardia.

Fuente: Elaboración propia con base en "What is AI bias?" (Holdsworth, 2025).

3. Revisión sistemática de casos en donde se aplica IA en la selección de personal y los factores para tener en cuenta a fin de reducir el impacto de los sesgos en el uso de la IA

A partir de la revisión sistemática de literatura, las principales evidencias que presenta el uso de IA en la selección de personal en cuanto incidencia de sesgos aparece con mayor frecuencia el *sesgo de género*. Morffi Collado y Bravo Places (2024) examinan la inadecuación de los marcos legales frente a la integración creciente de herramientas informáticas en empresas, con un enfoque específico en los procesos de selección de personal mediante inteligencia artificial (IA). A su vez, evidencian la necesidad de establecer límites a la libertad empresarial en el uso de tecnologías que podrían perpetuar sesgos y estereotipos de género.

En un segundo estudio, Quiroz Serrato y Noronha Ramos (2025) abordan los *sesgos algorítmicos* —en particular hacia mujeres embarazadas y personas con discapacidad— debido a un uso inadecuado de la IA y sin una legislación adecuada que regule su aplicación, lo que puede generar que los algoritmos sean programados con sesgos que resulten discriminatorios e impidan el acceso al empleo de determinados grupos de personas como mujeres, embarazadas y personas con discapacidad. Continuando con los *sesgos algorítmicos*, Delgado López (2024) examina la potencial vulneración del derecho a la no discriminación e igualdad de oportunidades que, al

igual que otros derechos fundamentales, puede verse con frecuencia amenazado por la opacidad de los algoritmos, la falta de transparencia en su funcionamiento o, simplemente, por la existencia de sesgos inherentes a los datos utilizados para entrenarlos.

Y, finalmente, Coral Lanchimba et al. (2025) destacan el impacto de la inteligencia artificial que se usa habitualmente en las empresas cuando se trata de procesos de selección de personal, en tanto promete optimizar recursos como es el tiempo de las personas que se encargan de los procesos, pero advierten que la dependencia de la IA puede debilitar habilidades humanas esenciales como la inteligencia emocional y el juicio crítico, lo que dificulta la evaluación de aspectos subjetivos como la empatía, vocación o compromiso. Dicho trabajo, orientado en entender las percepciones, las emociones y los significados en torno al uso de la IA mediante la investigación de tipo documental y descriptiva, basada en el análisis de fuentes secundarias como artículos científicos, tesis e informes relevantes entre 2017 y 2024, reveló los sesgos algorítmicos, la exclusión y deshumanización, y dejó en evidencia cómo estas tecnologías afectan y vulneran derechos laborales y humanos.

Conclusiones

Con base en lo expuesto, y a la literatura de casos relevada, la problemática de los sesgos en la selección de personal mediante aplicaciones de IA se concen-

tra en sesgos de género y algorítmicos, lo cual da cuenta de las dificultades que presenta el uso de estas tecnologías sin la adecuada preparación de los sistemas, cultura organizacional, ética y gobernanza de la IA. En cuanto a su reducción y prevención —ya que conforme pase el tiempo el uso de IA en procesos de selección de personal será más intensivo— será necesario que las organizaciones apliquen los principios evidenciados en la Tabla 2 de este trabajo para asegurarse de cumplir con la adecuada revisión de carga de datos en el modelo; dotar al equipo de RR. HH. de un *staff* variado e inclusivo, realizar una revisión constante del procesamiento de datos, asegurar el monitoreo constante de datos, y la procedencia e interoperabilidad de los datos recolectados de distinta infraestructura. Todas estas prevenciones a nivel organizacional son complementarias y acompañan a los procesos de adecuación legal que en materia de derecho laboral es necesario garantizar para un uso ético y legalmente adecuado de las tecnologías de IA en la selección de personal.

Referencias Bibliográficas

- Chiavenato, I. (2011). *Administración de Recursos Humanos. El capital humano de la organización* (9a ed.). McGraw Hill.
- Choi, A. (2021). España ante la Revolución Industrial 4.0: mercado laboral y formación. *Araucaria*, 47, 479-505. <https://shs.cairn.info/revista-araucaria-2021-2-page-479?lang=es&tab=texte-integral>
- CoralLanchimba, A. V., Gualotuña Quishpe, G. A. y Alarcón Toabanda, P. E. (2025). Desventajas del uso de la Inteligencia Artificial en los procesos de selección del personal: Disadvantages of using Artificial Intelligence in personnel selection processes. *Revista Multidisciplinar De Estudios Generales*, 4(3), 163-173. <https://doi.org/10.70577/reg.v4i3.165>
- Delgado López, I. (2024). El impacto de los sistemas algorítmicos en los procesos de selección de personal. Análisis jurídico-laboral a la luz del nuevo Reglamento europeo en materia de inteligencia artificial. *Revista De Trabajo Y Seguridad Social. CEF*, (483), 49-82. <https://doi.org/10.51302/rtss.2024.22257>
- Holdsworth, J. (2025). *What is AI bias?*. IBM. <https://www.ibm.com/>

- think/topics/ai-bias
- Manterola, C. y Otzen, T. (2015). Los sesgos en investigación clínica. *International Journal of Morphology*, 33(3), 1156-1164. <https://dx.doi.org/10.4067/S0717-95022015000300056>
- Márquez, I. (2017). El smartphone como metamedio. *Observatorio (OBS*)*, 11(2), 61-71. <https://10.15847/obsOBS11220171033>
- Monasterio Astobiza, A. (2021). Gobernanza global de la IA. *Cuadernos Salmantinos de filosofía*, 48, 187-211. <https://doi.org/10.36576/summa.144498>
- Morffi Collado, C. L. y Bravo Placeres, I. (2024). Sesgos de género en la selección laboral con inteligencia artificial: desafíos y soluciones. *GEDI-PRAXIS, Revista de Gestión, Educación y Ciencias Sociales*, 2(2), 100-128. <https://doi.org/10.5281/zenodo.14841783>
- Quiroz Serrato, C. A. y Noronha Ramos, A. J. (2025). El impacto discriminatorio de los sesgos algorítmicos en los procesos de selección de personal con discapacidad. *Laborem*, 24(31), 47-68. <https://doi.org/10.56932/laborem.24.31.2>
- Ramírez Autrán, R. (2023). Sesgos y discriminaciones sociales de los algoritmos en Inteligencia Artificial. Una revisión documental. *Entretextos*, 15(39), 1-17. <https://dialnet.unirioja.es/descarga/articulo/9380872.pdf>
- Valor Yébenes, J. A. (2024). Alan Turing y el origen de la inteligencia artificial: la superación de la intuición. *Naturaleza y libertad. Revista de estudios interdisciplinarios*, 18, 15-59. <https://dialnet.unirioja.es/descarga/articulo/9901836.pdf>