+IGUAL

NEGRO SOBRE BLANCO I

Impacto de los sesgos en Recursos Humanos a través de la IA

Enero 2024

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

NEGRO SOBRE BLANCO

(volumen 1)

EL IMPACTO DE LOS SESGOS EN RECURSOS HUMANOS A TRAVÉS DE LA IA

Índice

Introducción: por qué el Cámbrico influye en la gestión de talento 1

Estado de la cuestión: la universalidad de los sesgos 4

El impacto de los sesgos en la IAG aplicada a Recursos Humanos 7

Fundamentos de IA+Igual en el ámbito de los Recursos Humanos 12

Desafíos para Recursos Humanos 15

Conclusión 16

Bibliografía 19

Inteligencia Artificial

Introducción: por qué el Cámbrico influye en la gestión de talento

Todo está relacionado. Nada es casualidad. Lo que somos hoy como sociedad, la forma en que vivimos es el resultado de millones de cambios evolutivos, transformaciones y decisiones que, en muchos casos, se han ido produciendo de manera aleatoria y que nos ha hecho como somos. Entender esta realidad, profundizar en ella y analizarla, nos ayuda a conocernos mejor, a entendernos

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

a fondo. Y esa es la clave para poder ejercer, de la mejor manera posible, el escaso control que tenemos sobre nosotros mismos y lo que nos rodea.

Si ha leído hasta aquí, y le parece que todo esto no tiene nada que ver con usted, con su organización y, mucho menos, con la función de Recursos Humanos, se equivoca. Estamos hablando de personas, de individuos y de cómo la herencia de cientos de miles de años que llevamos impresa en nuestro ADN condiciona la forma en que entendemos la realidad y, por tanto, define cómo actuamos en nuestro día a día, dentro y fuera del trabajo. Y lo más importante para este análisis: cómo lo que somos se ha transmitido siempre a lo que hacemos, y hoy, más que nunca, a la Inteligencia Artificial.

Empecemos este viaje hacia el entendimiento en el periodo Cámbrico, hace unos 500.000 años. En aquel momento se produjo el estallido de vida más intenso jamás conocido. La explosión cámbrica dio lugar a la aparición de una increíble diversidad de vida sobre la tierra que incluye muchos de los principales grupos de animales presentes en la actualidad y que conocemos gracias al gran número de fósiles encontrados por todo el mundo, y más concretamente en China, Canadá y Groenlandia.

Desde el momento en el que se genera vida diversa, aparece la identidad como herramienta de defensa primaria: lo que soy, frente a lo que no soy; lo que como, frente a lo que me come. Este hecho no es baladí. Durante los posteriores miles de años en los que la evolución siguió su implacable camino, la "identidad" se mantuvo impresa en el ADN primigenio de los anfibios, los reptiles, las aves y los mamíferos.

Pensemos ahora en el cerebro. Ese complejo órgano de control es capaz de gestionar al mismo tiempo las funciones vitales, tomar decisiones o calcular riesgos mientras procesa los miles de estímulos por segundo que recibe desde los cinco sentidos. Su único objetivo: mantenerse con vida. Desde el más

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

sencillo, el reptiliano, al completo sistema nervioso central del ser humano, el cerebro está programado para sobrevivir. En su sofisticación, fue ganando complejidad y nuestro cerebro es, todavía hoy, una gran caja negra de la que no hemos podido extraer muchos de los secretos de su funcionamiento.

Pero algunas cosas sí sabemos. Por ejemplo, que un recurso evolutivo del cerebro es la capacidad de "simplificar" la información para detectar los peligros lo más rápido posible. Para ello, selecciona la información buscando todo aquello que le es "familiar" y lo cataloga como "peligroso" o "no peligroso". Es lo que, hace miles de años, nos ayudaba a identificar rápidamente a un "diente de sable" para salir corriendo. Este funcionamiento instintivo tiene efectos secundarios, como la "pareidolia".

Técnicamente sucede cuando el cerebro adjudica una identidad conocida a cualquier objeto, área o paisaje de la naturaleza. Lo más reconocible, puesto que vivimos en sociedad, es el rostro humano. Es por esto que podemos ver en formas abstractas algo parecido a un rostro: ojos, nariz, boca… Pensemos, por ejemplo, en el pomo de la puerta que abría Alicia para entrar en el país de las maravillas en la famosa película de Disney. En el *film*, la pareidolia toma vida: el pomo es la nariz; el ojo de la cerradura, la boca; y los tornillos que sujeta la manilla en la parte superior, los ojos.

Nuestro cerebro busca la forma de interpretar la información y ajustarla a lo que conoce y, hoy en día, que no hay depredadores salvajes acechando, su principal función consiste en reconocer caras y personas, con el añadido de identificar con prontitud las emociones que están manifestando por su expresión facial y gestos y que puedan revelarnos su intención hacia nosotros.

En esta herramienta de supervivencia biológica subyacen los sesgos cognitivos. El término "sesgo" se ha convertido en noticia por el impacto que tiene en el funcionamiento de la moderna Inteligencia Artificial (IA). ¿Podríamos estar

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

viviendo un nuevo Cámbrico, un momento de explosión tecnológica en la que surgen a diario nuevas aplicaciones de la Inteligencia Artificial? Aparquemos, de momento, la respuesta y hablemos de sesgos.

Estado de la cuestión: 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.

Existen numerosas relaciones de tipos de sesgos, pero en el ámbito que nos ocupa, el del mercado laboral y los recursos humanos, tenemos que poner el foco en los denominados "sesgos inconscientes" (*BIAS* en inglés). Los sesgos inconscientes son suposiciones, creencias o actitudes adquiridas que tenemos, pero no necesariamente somos conscientes de ellos. Si bien los sesgos son una parte normal de la función cerebral humana, a menudo pueden reforzar ciertos estereotipos produciendo más daño que beneficio a las empresas en lo que respecta a contratación, gestión de personas y toma de decisiones.

Los sesgos inconscientes se desarrollan con el tiempo a medida que acumulamos experiencias de vida y nos exponemos a diferentes estereotipos. Estos sesgos incluyen evaluaciones favorables y desfavorables, y se activan involuntariamente y sin la conciencia o el control intencional del individuo. Como resultado, influyen en nuestras creencias y comportamientos y, cuando

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

se trasladan a nuestra vida profesional, afectan a la forma en que contratamos, interactuamos y tomamos decisiones de negocios.
Identificamos 5 grandes grupos de sesgos inconscientes que impactan especialmente en las relaciones laborales.

1.- Sesgo de género. Es la asociación inconsciente de ciertos estereotipos con el género y la tendencia sexual. Este tipo de sesgo inconsciente afecta a las políticas de contratación, la dinámica de las relaciones en la empresa, las oportunidades laborales, el liderazgo o la política retributiva. En el ámbito de la RSC afecta a los objetivos de diversidad. Y desde el punto de vista normativo, se convierte en un problema legal por incumplimiento, en España por ejemplo, de la Ley de Igualdad. Hay que tener en cuenta, además, que el sesgo de género no es específicamente "femenino". En aquellas profesiones feminizadas (enfermería, cuidado infantil y de mayores por ejemplo) un algoritmo podría sesgar las candidaturas masculinas.

2.- Sesgo de edad. El edadismo se refiere a estereotipos en función de la edad. Este sesgo afecta a las relaciones laborales, la gestión del talento y las políticas de empleo. El principal impacto está en las políticas de contratación y en la rescisión de contratos; influye en la diversidad de las organizaciones y atenta contra los ODS, además de afectar a las políticas de RSC. En materia de empleo público, puede suponer un freno al acceso de personas mayores a programas de formación, lo que debilita la empleabilidad.

3.- Sesgo de apariencia. El *lookismo* se basa en un trato favorable hacia las personas que se ajustan a los prototipos de belleza establecido en cada cultura, lo que implica estatura, talla, color del cabello, capacidades diferentes. Afecta, directamente, a la decisión de contratación, lo que puede llegar a impactar en los resultados de negocio. También impacta en la forma en la que trabajamos en equipo, en las decisiones de promoción profesional y puede provocar problemas de acoso. Hay estudios que revelan, por ejemplo, que las personas

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

altas y rubias son potencialmente más "aceptables" en los procesos de contratación.

4.- Sesgo de intuición. Sucede cuando nos dejamos llevar por la primera impresión, ya sea buena o mala: construimos el "todo" de una persona con información limitada. Afecta a las decisiones de contratación y promoción, así como en la toma de decisiones a la hora de despedir personal. Este es el sesgo más subjetivo de todos y se basa en la experiencia y aprendizaje de cada individuo.

5.- Sesgo de confirmación. Selecciona y utiliza la información que confirme puntos de vista propios o expectativas. Todo aquello que es diferente a mí, es un potencial peligroso. Incide la raza, la cultura, la creencia religiosa, el origen, la clase social o la discapacidad. Puede afectar en la toma de decisiones de negocio porque todo aquello que es diferente, lejano o desconocido, produce un efecto inconsciente adverso que nos lleva a descartarlo. Internamente, produce fallos en la cultura de la organización.

Llegados a este punto, y antes de seguir avanzando en nuestro análisis, es importante diferenciar entre **prejuicio**, **estereotipo** y **sesgo.**

- Un prejuicio es una opinión generalmente negativa hacia una persona o grupo, que se forma sin motivo, sin el conocimiento necesario, y genera una actitud hostil y consciente que pretende actuar en consecuencia.

- Un estereotipo es la imagen simplificada y con pocos detalles a cerca de un grupo de personas que tienen en común ciertas características y que es asumido como cierto de forma colectiva por un grupo que, además, suele identificarse como "igual".

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

Prejuicios y estereotipos son la base de las actitudes discriminatorias que impactan en la convivencia y pueden suponer el maltrato y la limitación de derechos.

Los sesgos subyacen, de manera mucho más profunda, en nuestro subconsciente y, por tanto, impactan en nuestras acciones de manera inconsciente y no premeditada. Además, no nos identifican con ningún grupo y puede provocar actitudes hostiles o cordiales.

El impacto de los sesgos en la Inteligencia Artificial Generativa (IAG) aplicada a Recursos Humanos

Recopilando la tesis expresada hasta el momento, los sesgos son inherentes a la condición humana e influyen en nuestra toma de decisiones. Partiendo de esta premisa, ¿cómo impactan en el desarrollo de la IA? Estamos ante una tecnología basada en matemática avanzada computacional y estadística. Si añadimos la G de "generativa", la dotamos de capacidad de aprendizaje. La Inteligencia Artificial Generativa (IAG) se basa en modelos de aprendizaje automático para aprender los patrones y las relaciones de un conjunto de datos de contenido creado por personas. Y es en ese punto -"creado por personas"- donde está el quid de la cuestión, porque sus sesgos se transfieren de manera inconsciente al algoritmo.

El sesgo algorítmico se produce en aquellos casos en los que un determinado modelo algorítmico basado en datos produce repetidamente resultados no deseados por las personas que desarrollan, crean y entrenan el sistema. Con frecuencia, pero no siempre, esto se debe a que la recopilación y el uso de datos de entrenamiento están sesgados. En otras ocasiones, la causa son los problemas de interacción entre un algoritmo y otros procesos en un contexto concreto. Y en otras ocasiones, la falta de datos hace que el diseñador decida poner más peso en una variable por intuición, percepción del mercado o experiencia previa.

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

En la mayoría de estos casos, los sesgos se han incrustado en el modelo algorítmico de forma inconsciente. Como consecuencia, su funcionamiento produce resultados no deseados y provoca una forma de discriminación sistemática que afecta, generalmente, a los grupos sociales protegidos o vulnerables.

Pero, ¿cómo llegan ahí? Para entender el proceso, debemos tomar en consideración tres elementos básicos en el funcionamiento del desarrollo de la IAG.

1.- Los datos. Detrás de las herramientas de IAG, tal y como hemos visto, lo que hay son algoritmos de matemática compleja, computacional y estadística. Para conseguir las "predicciones" que se buscan, -por ejemplo, cuál es el mejor candidato para mi organización-, esos algoritmos deben trabajar con millones de datos. No miles, ni cientos de miles: millones. Estos datos son el quid de la cuestión; es en ese *big data* donde, de manera natural, la información está sesgada. Pensemos en una base de datos donde se recojan los datos profesionales de los mejores profesionales de enfermería: en el caso de España, por pura estadística, el perfil más eficiente sería el de una mujer, licenciada en enfermería y de origen español. Es decir: mujer blanca, heterosexual, católica. Aunque nos parezca curioso, si estos datos no se tienen en cuenta en la programación, todo aquello que sea "diferencial" a este prototipo definido como "excelente" será descartado.

Una multinacional del sector de bebidas, con expansión en todo el mundo, utiliza una herramienta de IAG a través de la cual se diseñan los itinerarios formativos de los empleaos de todas las filiales. La herramienta, subcontratada a un proveedor británico y que dice estar soportada en *machine learning*, está entrenada con 300.000 datos. Parece poco fiable el resultado que pueda dar una herramienta de este tipo, entrenada solo con 300.000 datos para un proyecto multi país.

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

2.- La globalidad. Si tenemos en cuenta que los desarrolladores de IAG aplicada a RR.HH. son principalmente estadounidenses, debemos pensar que, si la BB.DD. utilizada se corresponde al mercado de trabajo de ese país, muy posiblemente tendrá sesgos que, en España, supondrían incumplimientos normativos, por ejemplo. En materia de mercado laboral y recursos humanos, la información utilizada para su aplicación es especialmente delicada. La normativa laboral es diferente por países, así como, por ejemplo, la diversidad social y la protección de la discapacidad. La Ley de Protección de Datos es un elemento que influye también en esta cuestión, afectando al uso de información.

3.- Ajuste de modelos. En consecuencia, y si hay sensibilidad acerca de la posible incorrección de la base de datos, en ocasiones se decide dar relevancia en el modelo algorítmico a cuestiones definidas por los desarrolladores o la empresa que va a utilizar la herramienta. Este proceso, muy delicado, puede causar de nuevo problemas en el funcionamiento al incluir sesgos de manera inconsciente que desvirtúen el resultado del modelo.

Con todo, los procesos de recursos humanos y el mercado laboral en general son un entorno en el que la aplicación de la IAG serviría para impulsar y optimizar resultados. Imaginemos, por ejemplo, un sistema público de empleo que cuente con una herramienta algorítmica alimentada con datos de las personas en búsqueda activa de empleo (perfil, formación, experiencia, habilidades, tiempo en desempleo, prestaciones que reciben, etc.), ofertas en activo del mercado laboral (requisitos de experiencia, habilidades, convocatorias de empleo público, etc.) y la oferta formativa existente para la recapacitación profesional garantizase un funcionamiento óptimo del sistema "prediciendo" opciones de empleo óptimas para cada candidato.

Pero, ¿cómo garantizamos que una herramienta con este elevado potencial funciona correctamente? ¿Cómo definimos su modelo de aprendizaje para que incluya en el proceso a exreclusos, por ejemplo? ¿Cómo puede ser

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

entrenado para que vea solo el talento de cada persona y no se vea influido por sesgos relacionados con la discapacidad, el género o la raza?
El primer paso ha sido la estrategia europea sobre la IA que tiene como objetivo convertirse en referencia mundial de esta tecnología, con un enfoque humanista. En el último semestre de 2023, y durante la presidencia de turno de España del Consejo de la UE, se ha logrado alcanzar un acuerdo provisional sobre el reglamento europeo de uso de la IA centrada en salvaguardar los derechos de los ciudadanos, sin perder las oportunidades de desarrollo, innovación y competitividad. Tras tres días de conversaciones, la negociación se cerró con un acuerdo provisional sobre la propuesta de normas armonizadas sobre inteligencia artificial que, se espera, se convierta en Ley en 2026.

Esta futura Ley de IA es una iniciativa pionera en el mundo que busca conseguir el equilibrio entre el potencial desarrollo de una IA segura y confiable en el mercado único por parte de los sectores público y privado y estimular, al mismo tiempo, la inversión y la innovación en IA en Europa.

La prioridad del Parlamento Europeo es garantizar que los sistemas de IA utilizados en la Unión Europea sean seguros, transparentes, trazables, no discriminatorios y respetuosos con el medio ambiente. El reto es establecer un marco legal que encuentre el equilibrio entre esta garantía de uso ético, seguro y confiable de la IA y el impulso al desarrollo, la innovación y la competitividad que supone su uso. Sobre todo, teniendo en cuenta que fuera de la UE la normalización del uso de la IA no está en la hoja de ruta de países como EE.UU., China o India y, por tanto, su desarrollo y aplicación está muy lejos de garantizar el objetivo marcado por la UE.

En nuestro caso de estudio, lo más importante de este borrador es la clasificación de la IA en cuatro niveles de riesgo:

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

1.- Riesgo inaceptable. Pertenecen a esta clasificación todos aquellos sistemas de inteligencia artificial cuyo uso implique un riesgo inadmisible para la seguridad, la vida y los derechos fundamentales. Son, por tanto, desarrollos PROHIBIDOS en el marco de la UE. Incluye sistemas capaces de manipular el comportamiento humano, clasificar a las personas en función de su comportamiento, estatus socioeconómico o características, o discriminan a través de sistemas de identificación biométrica en tiempo real y a distancia, como el reconocimiento facial.

2.- Riesgo Alto. Se incluyen en este grupo todos los sistemas que afecten negativamente a la seguridad o a los derechos fundamentales, y se dividirán en dos categorías:

- Los sistemas de IA que se utilicen en productos sujetos a la legislación europea sobre seguridad de productos. Incluye juguetes, aviación, automóviles, dispositivos médicos y ascensores.
- Los sistemas de IA pertenecientes a ocho ámbitos específicos que deberán registrarse en una base de datos de la UE:
  - ✓ Identificación biométrica y categorización de personas físicas
  - ✓ Gestión y explotación de infraestructuras críticas
  - ✓ Educación y formación profesional
  - ✓ Empleo, gestión de trabajadores y acceso al autoempleo
  - ✓ Acceso y disfrute de servicios privados esenciales y servicios y prestaciones públicas
  - ✓ Aplicación de la ley
  - ✓ Gestión de la migración, el asilo y el control de fronteras
  - ✓ Asistencia en la interpretación jurídica y aplicación de la ley

Todos los sistemas clasificados de "alto riesgo" serán evaluados antes de su comercialización y a lo largo de su ciclo de vida. De esta recomendación europea nace el proyecto IA+Igual, que propone un modelo de análisis

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

algorítmico de herramientas de IA aplicadas a los procesos de recursos humanos y el mercado laboral.

3.- Riesgo Limitado. Son aquellos que no suponen un alto riesgo para los derechos y las libertades. Deben ser transparentes y los usuarios deben ser conscientes de que están interactuando con una máquina para que puedan tomar una decisión informada sobre continuar o retroceder, como en el caso de los *chatbots*.

4.- Riesgo Mínimo. Serán todos aquellos que no suponen implicaciones para los derechos de los usuarios, por lo que no encajan en el resto de las clasificaciones. Unos ejemplos serían los filtros de spam o la IA aplicada a videojuegos. La gran mayoría de los sistemas de IA utilizados actualmente en la UE entran en esta categoría.

Como hemos visto, todas las herramientas de IA aplicadas a procesos de recursos humanos y del mercado laboral se consideran de alto riesgo. El impacto que tienen los sesgos en la IA aplicada a estos ámbitos supone un retroceso en los avances conseguidos en materia de igualdad de derechos y gestión de la diversidad.

Además, supone la vulneración de derechos fundamentales de las personas porque atenta contra los ODS en materia de Igualdad de Género (5); Trabajo Decente y Crecimiento Económico (8); y Reducción de las Desigualdades (10). Y, por otro lado, supone el incumplimiento grave de la normativa laboral, que puede tener consecuencias civiles y penales para los infractores.

Fundamentos de IA+Igual en el ámbito de los Recursos Humanos

En 2015, Amazon implantó una herramienta, basada en IA, para la selección de personal especializado en software y otras áreas técnicas: debido a determinados sesgos erróneos ("cajas negras") de su algoritmo, prefería y elegía

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

personas del sexo masculino en detrimento del sexo femenino. En 2018, el sistema de IAG de Google identificaba a los hombres de raza negra como gorilas.
En 2020, una sentencia de un tribunal italiano determinó que el algoritmo de la plataforma de gestión de repartos "Frank", utilizada por Deliveroo, penalizaba el absentismo de los repartidores sin diferenciar los incumplimientos por motivos baladíes respecto de las ausencias por enfermedad, minusvalía, cuidado de menores o para ejercer su derecho de huelga (supuestos, estos últimos; legítimos y derechos fundamentales de los ciudadanos, los otros).

Son solo tres ejemplos, pero la realidad demuestra que decisiones que antes eran tomadas por humanos, hoy son tomadas por algoritmos de IAG, como las relativas a la contratación de personas o la evaluación del desempeño.

Tal y como hemos explicado, el uso de estos sistemas conlleva un riesgo, ya que los datos con los que los algoritmos son entrenados están condicionados por nuestros conocimientos y sesgos. Por este motivo, para evitar que un algoritmo discrimine a ciertos colectivos, es necesario verificar que sus datos de aprendizaje no contienen ningún sesgo.

La RAE define **algoritmo** como un código software que procesa un conjunto limitado de instrucciones. Para crear un algoritmo es necesario un proceso estadístico, matemático y humano muy complejo, que incluye recopilación, preparación y análisis de datos en diversas etapas y en el que influyen las decisiones de desarrolladores y directivos. Los algoritmos pueden incluir sesgos denominados **inconscientes** debidos a malas elecciones de datos, datos sesgados contra un grupo, pobres, incompletos, incorrectos, desactualizados, mal recopilados o de mala calidad.

Los **sesgos inconscientes** en la IAG aplicada a herramientas que se utilizan en los procesos de recursos humanos -como selección de personal (catalogación

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

de CV, por ejemplo), evaluación de desempeño, carrera profesional, predicción de fuga de talento, etc.- han sido catalogados por la UE como de alto riesgo. Su uso inadecuado, tiene un impacto negativo en las políticas de igualdad de oportunidades, gestión de la diversidad y accesibilidad al empleo de los estados miembro, y vulneran los derechos fundamentales de las personas, especialmente las que pertenecen a los colectivos más desfavorecidos.

Esto es porque los sesgos inconscientes modelan el aprendizaje de los algoritmos y terminan discriminando a las personas candidatas, desde el inicio de los procesos de selección, por su sexo, origen, edad, procedencia o discapacidad. Todas estas cuestiones nada tienen que ver con el talento y, además, de suponer la vulneración de los derechos de las personas, impiden a las empresas la posibilidad de incorporar o promocionar el mejor talento, lo que a medio plazo impacta en el resultado del negocio.

Si determinamos que los sesgos en la IAG aplicada a los procesos de recursos humanos son una cuestión que -además de afectar a los resultados de negocio- puede suponer incumplimientos graves de la normativa laboral, y añadimos el hecho de que ya han sido identificados por la UE como de alto riesgo, concluimos que es imperativo poner en marcha proyectos dirigidos a encontrar soluciones eficientes a esta realidad.

Una de las líneas de trabajo es la creación de un modelo de normalización y certificación que permita auditar estas herramientas, identificar los potenciales sesgos e instrumentar su eliminación del algoritmo. Llevar a cabo una prueba empírica de cómo están funcionando los algoritmos nos permite extraer una radiografía de la realidad para trabajar sobre la experiencia y no sobre intuiciones.

El proyecto IA+Igual, financiado con fondos *Next Generation* a través del plan de Transformación y Resiliencia del Gobierno de España, es una iniciativa de

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

innovación social de la Comunidad de Madrid que está llevando a cabo la auditoría de IA aplicada a procesos de recursos humanos. Las primeras evidencias extraídas del ámbito de investigación revelan un enorme desconocimiento tecnológico por parte de los responsables de RR.HH. de las organizaciones, un desinterés por invertir en la mejora de procesos en este ámbito -amparado en el riesgo reputacional que supone- y la necesidad de impulsar la comunicación interna cuando se aplican estas herramientas para contrarrestar la imagen negativa que la IA tiene en la sociedad.

Afrontar el desconocimiento, la desinformación y el miedo son tres de los retos básicos que hay que acometer para promover la implantación de IAG en el ámbito laboral. Cuestión que, sin duda, conseguiría optimizar recursos, gestionar de una manera más eficiente el talento y podría significar el avance definitivo en materia de igualdad de oportunidades.

Desafíos para Recursos Humanos

Recursos humanos se enfrenta al reto de utilizar en sus procesos herramientas de IAG que optimicen sus resultados. La competitividad de las organizaciones pasa por una gestión del talento eficiente que consiga atraer a los profesionales del mercado global y retenerlos con sistemas de evaluación, promoción y retribución de última generación. La experiencia-empleado se ha convertido en la clave del éxito para muchas organizaciones y la IAG aplicada proporciona la posibilidad de analizar datos y generar información de alto valor para conseguirla.

¿Por qué, entonces, solo un 14% de los profesionales del área trabajan con esta tecnología? Mientras que el 76% considera que la IAG será fundamental en el futuro, son pocas las organizaciones que se han lanzado a introducirla en sus procesos de RR.HH.; mientras sí lo hacen, con buenos resultados, en otros procesos de negocio. Los desafíos son claros:

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

- Impacto reputacional. El uso de IAG en procesos de RR.HH., sin la garantía de un funcionamiento ético y confiable, arrastra el riesgo de demandas laborales por incumplimientos normativos relacionados con la igualdad y la diversidad. Llegado el caso, podrían impactar negativamente en el valor reputacional de la marca. Como alternativa, las empresas se deciden por la subcontratación para evitar responsabilidades civiles y penales, pero, con ello, no se salva el impacto en la reputación.
- Ampliar el conocimiento. Los departamentos de recursos humanos no cuentan con perfiles tecnológicos que supervisen la adquisición y funcionamiento de herramientas de IAG aplicadas a sus procesos. Cada herramienta debe ser entrenada con datos apropiados y adaptados al objetivo de la compañía y a sus perfiles profesionales. Para ello, deberían contar con consejos asesores multidisciplinares que supervisaran la selección de datos con los que se entrena el algoritmo y el diseño del modelo de aprendizaje.
- Comunicación focalizada. La controversia generada por las informaciones difundidas por los medios de comunicación, provocan rechazo por parte de la ciudadanía. El miedo "irracional" a ser despedido por un algoritmo subyace en las organizaciones, y la implantación de estas herramientas de IAG debe ir acompañada por una campaña de información y sensibilización adecuada, que garantice su usabilidad.
- Garantía normativa. La legislación laboral debe acompañar, además, al marco de uso de esta tecnología. A pesar de nuevas normas, como la *UNE19602 de Compliance Laboral* aprobada en julio de este mismo año, todavía queda mucho por hacer.

Conclusión

En las últimas dos décadas se ha avanzado mucho en el terreno de la igualdad y la diversidad en el ámbito laboral. Se han desarrollado normativas que impulsan el avance hacia la igualdad de oportunidades, con modelos de organización que apuestan por la inclusión y la diversidad de todos los

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

colectivos, reconociendo el valor del talento diverso y las capacidades diferentes. Todos esos avances que se han realizado, no solo desde el ámbito legislativo y de la Administración, sino también desde las organizaciones a través de los planes de Responsabilidad Social Corporativa, el cumplimiento de los Objetivos de Desarrollo Sostenible (ODS) y la Agenda 2030, y que se refrenda, cada año, con un aumento del número de empresas que presentan su Memoria No Financiera, corren el riesgo de perderse si no establecemos sistemas de supervisión que garanticen el funcionamiento ético y confiable de la IAG.

No cabe duda de que la IAG ha llegado para quedarse. Que pronto se convertirá en herramienta imprescindible en el ámbito de los RR.HH., y que las empresas solo podrán ser competitivas si son capaces de atraer al mejor talento. Esté donde esté.

Para ello, y siguiendo las recomendaciones de la UE que, previsiblemente, se convertirán en Ley en 2026, es imprescindible establecer un modelo de trabajo que permita realizar, de manera sistemática, la auditoría y certificación de los algoritmos aplicados en IAG, empezando por aquellos que ya han sido identificados como de **alto riesgo.** Debe ser un modelo de certificación algorítmica escalable y personalizado por cada uno de los ámbitos de uso de la IA. Y debe ser realizado por entidades certificadoras externas que trabajen con estándares éticos y confiables internacionales.

Los procesos de RR.HH. y los relacionados con el mercado laboral son un ámbito prioritario sobre el que establecer estos mecanismos de supervisión, análisis y control algorítmico. Y los sesgos representan el foco de trabajo sobre el que debemos poner el foco. El objetivo es doble: por un lado, promover que las empresas contraten el mejor talento para impulsar su competitividad; y, por otro, garantizar la igualdad de oportunidades a todos los colectivos, incluso los identificados como "vulnerables".

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

La IAG no tiene conciencia. No sabe lo que es verdad o lo que no. No sabe lo que está bien o lo que está mal. Y no se cuestiona así misma por los resultados que obtiene al analizar los datos. Todavía no, al menos. Por eso, es esencial establecer sistemas de trabajo "humanistas" en torno a los procesos puramente "técnicos". El proyecto IA+Igual tiene como objetivo crear una IAG integradora que garantice que el acceso al mercado de trabajo se gestiona de una forma equitativa y justa; y en función únicamente del talento de la persona candidata y su adecuación al puesto. La plataforma de gestión IA+Igual permite configurar el modelo de análisis para analizar los sesgos inconscientes de los algoritmos no solo desde la perspectiva de género, sino también de discriminación por edad, orientación sexual, grado de discapacidad, procedencia, religión, etc.

Para ello, además de tecnología, es necesario innovar en los procesos de toma de decisiones a la hora de adquirir esta tecnología; actualizar los conocimientos que los departamentos de RR.HH. tienen sobre la materia; formar y sensibilizar a los profesionales acerca de las posibles repercusiones negativas que tiene su trabajo a causa de estos sesgos inconscientes; y acompañar a la sociedad en su propio proceso de digitalización, perdiendo el miedo al uso de la IA, pero comprendiendo los riesgos que supone su uso inapropiado.

Hasta ahora, hemos dejado el desarrollo de la tecnología en manos de técnicos: el resultado ha sido una tecnología cada vez más avanzada, sofisticada y eficiente. En este nuevo período cámbrico de explosión de nuevas formas de IAG, no se está teniendo en cuenta la transmisión que estamos haciendo de nuestro "fallo humano" a su modelo de aprendizaje. Ha llegado la hora de ponerlo **Negro sobre Blanco**.

Marisa Cruzado
CEO de IA+Igual

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

Bibliografía

Díaz Mairal, Francisco (2023): *Cultura del dato, inteligencia artificial y RRHH*. Capital Humano, 386.

Fuller, P., Chow, A., Murphy, M. (2023): *Sesgos inconscientes: cómo reformularlos, cultivar conexiones y crear equipos de alto rendimiento*. EE.UU.: Conecta

Raghavan, M., Barocas, S., Kleinberg, J., Levy, K. (2020): *Mitigating bias in algorithmic hiring: evaluating claims and practices*. FAT* 2020: pp. 469-481. https://doi.org/10.1145/3351095.3372828

Rueda, Francisco J. (2023): *Retos pendientes en el uso de la Inteligencia Artificial en el sector de los Recursos Humanos*. CEPAL.

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