Fundamentos algorítmicos y ética en IA: de la teoría a la práctica

Kit de herramientas para sesiones sincrónicas

UC5 | Casos de estudio y proyectos

Soporte de diapositivas de PowerPoint

Número de proyecto | 2022-1-ES01-KA220-HED-000085257

ÍNDICE

- INTRODUCCIÓN – 3
- SESGOS ALGORÍTMICOS Y DE APRENDIZAJE AUTOMÁTICO – 6
- SESGO ALGORÍTMICO Y SUS IMPLICACIONES EN EL MUNDO REAL - EJEMPLOS DE CASOS - 16
- HERRAMIENTAS PARA DETECTAR SESGOS ALGORÍTMICOS – 28
- PROYECTO | IA EN RECLUTAMIENTO - 35

INTRODUCCIÓN

FUENTE DE LA IMAGEN | Freepik

EN ESTA UNIDAD DE COMPETENCIA ENCONTRARÁS: LAS SIGUIENTES MATERIAS:

•Presentación de diversos sesgos algorítmicos y de aprendizaje automático a través de estudios de casos relacionados con TikTok y tecnologías de reconocimiento facial.

•Una visión general de las herramientas diseñadas para identificar sesgos algorítmicos y sus consecuencias.

•Un proyecto en profundidad sobre el uso de la IA en la contratación, que pone de manifiesto los sesgos en varias fases del proyecto.

•El papel de los enfoques conscientes de la equidad en el aprendizaje automático y la implementación de procesos de auditoría de sesgos.

AL FINAL DE LA UNIDAD DE COMPETENCIA, USTED DEBERÍA SER CAPAZ DE:

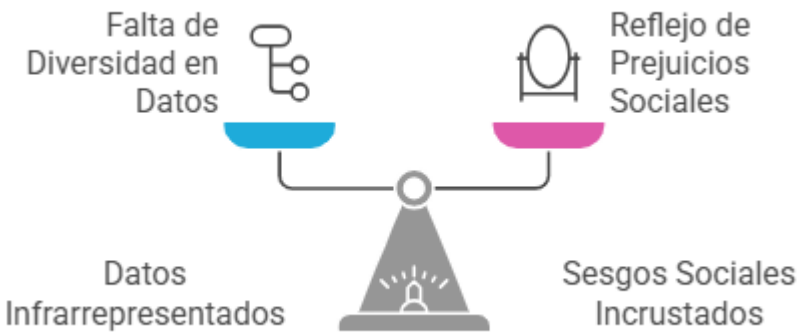
- Identificar diferentes tipos y causas de sesgo algorítmico, y ejemplos de tipos de sesgo en conjuntos de datos de aprendizaje automático
- Evaluar el sesgo algorítmico y sus implicaciones en el mundo real
- Analizar diferentes herramientas para detectar sesgos algorítmicos
- Analizar las fases del proyecto – Caso "IA en reclutamiento"

SESGOS ALGORÍTMICOS Y DE APRENDIZAJE AUTOMÁTICO

FUENTE DE LA IMAGEN | Freepik

Sesgos algorítmicos y de aprendizaje automático

¿QUÉ ES EL SESGO ALGORÍTMICO Y ¿POR QUÉ ES IMPORTANTE TENERLO EN CUENTA?



FUENTE DE LA IMAGEN | NAPKIN.AI

- En el aprendizaje automático, los algoritmos utilizan conjuntos de datos o datos de entrenamiento para aprender patrones.
- El sesgo algorítmico ocurre cuando el algoritmo toma decisiones que sistemáticamente perjudican o discriminan a ciertos grupos de personas, por ejemplo, en función de la raza o el género.
- El sesgo algorítmico puede aparecer si un determinado grupo de personas está infrarrepresentado en los datos o si los sesgos sociales existentes están incrustados en los propios datos.
- Es crucial comprender las causas e implicaciones del sesgo algorítmico para desarrollar soluciones de IA éticas y confiables.

(Harvard Business Review, S. Friis y J. Riley, 2023)

Sesgos algorítmicos y de aprendizaje automático

Diferentes tipos de sesgos algorítmicos

Los sesgos algorítmicos se pueden dividir, por ejemplo, en tres categorías:

1. **Sesgo preexistente. Sesgo en un sistema que existe de forma independiente y, por lo general, antes de la creación del sistema. Pueden entrar en un sistema de forma intencionada o no intencionada, ya sea a través de esfuerzos explícitos o acciones inconscientes, incluso con buenas intenciones.**

2. **El sesgo técnico surge de restricciones técnicas o consideraciones técnicas. Las fuentes de sesgo técnico se pueden encontrar en varios aspectos del proceso de diseño.**

3. **El sesgo emergente surge en un contexto de uso con usuarios reales. Este sesgo suele surgir en algún momento después de que se completa un diseño, como resultado de un cambio en el conocimiento de la sociedad, la población o los valores culturales.**

Friedman y Nissenbaum (1996)

Sesgos algorítmicos y de aprendizaje automático

Diferentes causas de sesgo

FUENTE DE LA IMAGEN | Imagen rediseñada (la imagen original: British Medical Journal)

Las tecnologías digitales tienden a reforzar las desigualdades que pueden aparecer, por ejemplo, como aplicaciones discriminatorias. Esta es una causa de discriminación en la toma de decisiones de la IA, que resulta de las brechas éticas en el desarrollo, la implementación o la disponibilidad de estas tecnologías.

Sesgos algorítmicos y de aprendizaje automático

Ejemplo de tipos de sesgo en Conjuntos de datos de aprendizaje automático

Los ingenieros entrenan modelos de aprendizaje automático alimentándolos con un conjunto de ejemplos de entrenamiento. La participación de los humanos en la selección y entrega de los datos puede causar sesgos.

Analicemos los siguientes tipos de sesgos:

- Sesgo en la presentación de informes
- Sesgo de atribución grupal
- Sesgo implícito
- Sesgo de automatización
- Sesgo de selección

Google para desarrolladores; Aprendizaje automático; Equidad: Tipos de sesgo

Sesgos algorítmicos y de aprendizaje automático

Ejemplo de tipos de sesgo en Conjuntos de datos de aprendizaje automático

Sesgo en la presentación de informes

- La frecuencia de los eventos, las propiedades o los resultados de un conjunto de datos no refleja la frecuencia del mundo real.
- El sesgo surge de un enfoque en documentar circunstancias inusuales o memorables.
- Las situaciones ordinarias pueden ser descuidadas, ya que se supone que son evidentes por sí mismas.

Ejemplo

En un entrenamiento de modelo de análisis de sentimientos, predice si las reseñas de libros son positivas o negativas en función de los envíos de los usuarios a un sitio web popular. La mayoría de las reseñas en los datos de entrenamiento expresan opiniones extremas (ya sea que amen u odien un libro) porque es menos probable que las personas revisen un libro si tienen una reacción moderada. Como resultado, el modelo tiene dificultades para predecir correctamente el sentimiento de las reseñas utilizando un lenguaje sutil para describir un libro.

Google para desarrolladores; Aprendizaje automático; Equidad: Tipos de sesgo

Sesgos algorítmicos y de aprendizaje automático

Ejemplo de tipos de sesgo en Conjuntos de datos de aprendizaje automático

Sesgo de atribución grupal

El sesgo de atribución grupal es la tendencia a aplicar características individuales a todo un grupo al que pertenecen. Hay dos formas principales:

- **Sesgo dentro del grupo: preferir miembros de su propio grupo o características compartidas.**

  Ejemplo | Dos ingenieros que entrenan en un modelo de evaluación de currículos para desarrolladores de software pueden favorecer a los solicitantes de la misma academia de ciencias de la computación a la que asistieron, asumiendo que están más calificados.

- Sesgo de homogeneidad fuera del grupo: estereotipar a los individuos de un grupo al que no perteneces, ver sus rasgos como más uniformes.

  Ejemplo | Dos ingenieros que entrenan un modelo de selección de currículos pueden asumir que todos los solicitantes que no asistieron a una academia específica de ciencias de la computación carecen de experiencia para el puesto.

Google para desarrolladores; Aprendizaje automático; Equidad: Tipos de sesgo

Sesgos algorítmicos y de aprendizaje automático

Ejemplo de tipos de sesgo en Conjuntos de datos de aprendizaje automático

Sesgo implícito

El sesgo implícito ocurre cuando se hacen suposiciones basadas en modelos mentales personales y experiencias que pueden no aplicarse universalmente.

Ejemplo | Un ingeniero que entrena un modelo de reconocimiento de gestos utiliza un movimiento de cabeza para indicar "no". Sin embargo, en algunas regiones, un movimiento de cabeza significa "sí".

- Un tipo común de sesgo implícito es el sesgo de confirmación, en el que los creadores de modelos procesan inconscientemente los datos para afirmar las creencias existentes.
- A veces, los creadores de modelos pueden seguir entrenando hasta que los resultados se alineen con su hipótesis original, lo que se conoce como sesgo del experimentador.
    - Ejemplo | Un ingeniero que construye un modelo de agresividad canina asocia la hiperactividad con un encuentro infantil con un caniche toy. A pesar de que el modelo inicialmente retrata a los caniches toy como tranquilos, el ingeniero continúa reentrenándolo hasta que indica que los caniches más pequeños son más agresivos, afirmando así su sesgo inicial.

Google para desarrolladores; Aprendizaje automático; Equidad: Tipos de sesgo

Sesgos algorítmicos y de aprendizaje automático

Ejemplo de tipos de sesgo en Conjuntos de datos de aprendizaje automático

Sesgo de automatización

Preferencia por los resultados de los sistemas automatizados sobre los no automatizados, independientemente de sus respectivas tasas de error.

Ejemplo

Los ingenieros de software de un fabricante de engranajes estaban entusiasmados con la presentación de un modelo "innovador" recién entrenado diseñado para detectar engranajes defectuosos. Sin embargo, su entusiasmo se desvaneció cuando un gerente de fábrica señaló que las tasas de precisión y recuperación del modelo eran un 15% más bajas que las logradas por los inspectores humanos.

Google para desarrolladores; Aprendizaje automático; Equidad: Tipos de sesgo

Sesgos algorítmicos y de aprendizaje automático

Ejemplo de tipos de sesgo en Conjuntos de datos de aprendizaje automático

Sesgo de selección

El sesgo de selección se produce cuando los ejemplos de un conjunto de datos no se eligen de forma que representen su distribución en el mundo real. Puede tomar varias formas.

- **Sesgo de cobertura: los datos no se seleccionan de forma representativa.**

  Ejemplo | Un modelo de IA se entrena para recomendar restaurantes en función de las opiniones de los clientes. Sin embargo, el modelo solo se entrena con las reseñas de una cadena de restaurantes popular, ignorando las reseñas de otros restaurantes. Como resultado, es posible que las recomendaciones del modelo de IA no reflejen con precisión los diferentes gustos y preferencias de todos los clientes potenciales, ya que no ha tenido en cuenta las opiniones de los clientes que visitaron otros restaurantes.

- **Sesgo de no respuesta (o sesgo de participación): los datos dejan de ser representativos debido a las brechas en la participación durante la recopilación de datos.**

  Ejemplo | Un sistema de recomendación de películas basado en IA recopila comentarios de los usuarios de una plataforma de transmisión en particular, pero ignora los comentarios de los usuarios de otras plataformas, lo que hace que sus preferencias no estén adecuadamente representadas. Como resultado, es posible que las recomendaciones de películas generadas por IA no reflejen eficazmente los diferentes gustos de todos los usuarios debido a la falta de comentarios de los usuarios de las diferentes plataformas de transmisión.

  Sesgo de muestreo: no se utiliza una aleatorización adecuada durante la recopilación de datos.

- **Sesgo de muestreo: no se utiliza una aleatorización adecuada durante la recopilación de datos.**

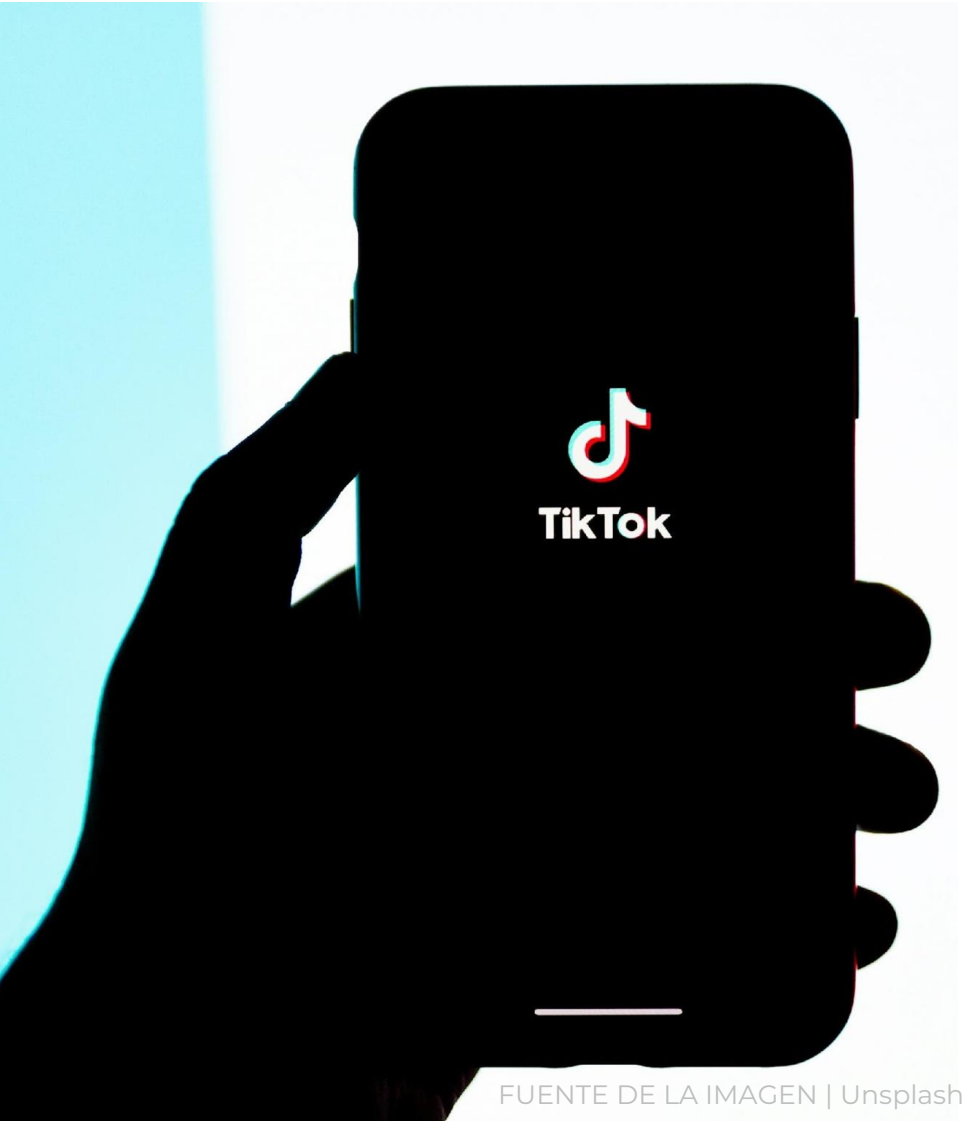
  Ejemplo | Un modelo de IA se entrena para predecir las ventas futuras de un nuevo producto encuestando a las personas que lo compraron y a las personas que compraron un producto similar sobre sus hábitos de compra. En lugar de elegir personas al azar, los encuestadores eligieron a las primeras 200 que respondieron a la encuesta, lo que podría significar que estaban más entusiasmados con el producto que la mayoría de los compradores.

SESGO ALGORÍTMICO Y SUS IMPLICACIONES EN EL MUNDO REAL: EJEMPLOS DE CASOS

FUENTE DE LA IMAGEN | Freepik

Sesgos algorítmicos y de aprendizaje automático

CASO – El algoritmo dañino de TikTok



FUENTE DE LA IMAGEN | Unsplash

La empresa nacional de medios de comunicación finlandesa Yle investigó qué tipo de contenido muestra el algoritmo de TikTok a una adolescente de 13 años que sufre depresión y una imagen corporal distorsionada.

- Yle creó un perfil para el personaje ficticio Ella que representa a la niña deprimida de 13 años.
- El científico de datos de Yle navegó por TikTok con el perfil de Ella durante poco más de cinco horas.

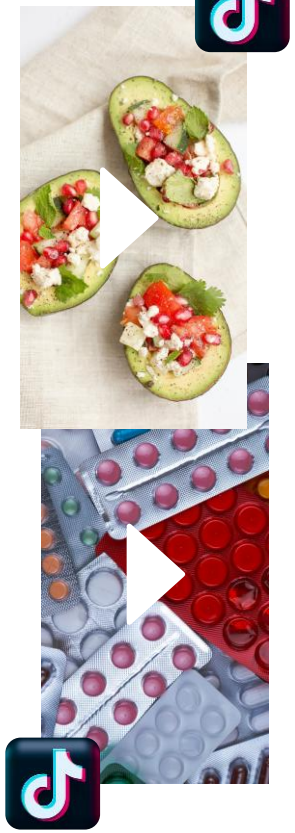
Echa un vistazo a los videos que TikTok mostró para Ella en el artículo de Yle

Yle, 11 de noviembre de 2023

Sesgos algorítmicos y de aprendizaje automático

CASO – El algoritmo dañino de TikTok

Los resultados fueron alarmantes



1. HORA - contenido nocivo 7% de 100%

(Los videos sobre problemas de salud mental y problemas alimentarios se contaron como contenido dañino).

- Al principio: videos alegres sobre comida, cachorros, gatos, patinetas y bromas.
- Ella mira los videos de comida un poco más.
- Después de solo 5 minutos, Ella se detiene para mirar un video sobre una mujer que está gritando.
- En solo un momento, TikTok muestra un video sobre antidepresivos.
- Más videos sobre salud mental y comidas ligeras.

"El algoritmo de Tiktok aprende rápidamente qué temas le interesan a un navegador".

Yle, 11 de noviembre de 2023

2. HOUR - harmful content **28%** out of 100%

- The videos' tone gets darker.
- Occasional content on subjects such as hair care and animals.
- **Content related to suicide and self-harm is now starting to appear in the videos.**
- Among the videos, there is also content that, for example, encourages people to keep going despite difficulties.

FUENTE DE LAS IMÁGENES | Unsplash y Pexels

Imágenes: Unsplash y Pexels

Sesgos algorítmicos y de aprendizaje automático

CASO – El algoritmo dañino de TikTok

Los resultados fueron alarmantes

"En un video, una mujer con un trastorno mental habla sobre cómo a veces siente ganas de quitarse la vida".

Yle, 11 de noviembre de 2023

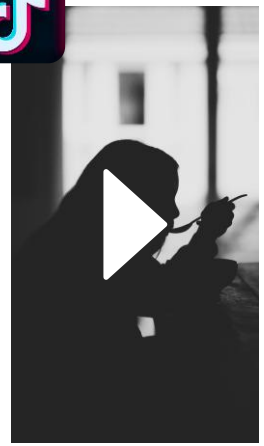
Sesgos algorítmicos y de aprendizaje automático

CASO – El algoritmo dañino de TikTok

Los resultados fueron alarmantes

3. HOUR - harmful content **68%** out of 100%

- **Just before the start of the third browsing hour, Tiktok shows Ella the first video on eating disorders.**
- After the first video, more eating disorder content starts to appear, and Ella stops to watch it.
- More content on low-calorie foods and self-harm.



4. HOUR - harmful content **57%** out of 100%

- Slightly fewer videos about depression and self-harm.
- Instead, the Tiktok algorithm seems to have recognised that Ella is particularly interested in light meals and the toxic content that encourages her to limit her eating.
- More food and weight loss videos than before.

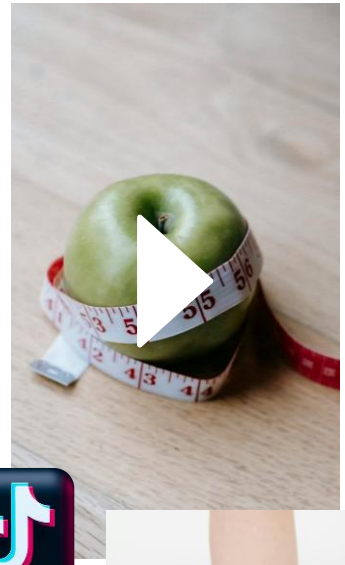
Yle, 11 de noviembre de 2023

FUENTE DE LAS IMÁGENES | Unsplash y Pexels

Sesgos algorítmicos y de aprendizaje automático

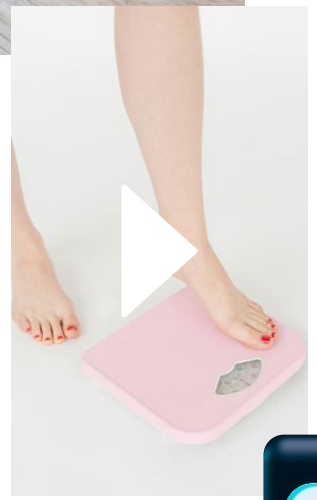
CASO – El algoritmo dañino de TikTok

Los resultados fueron alarmantes



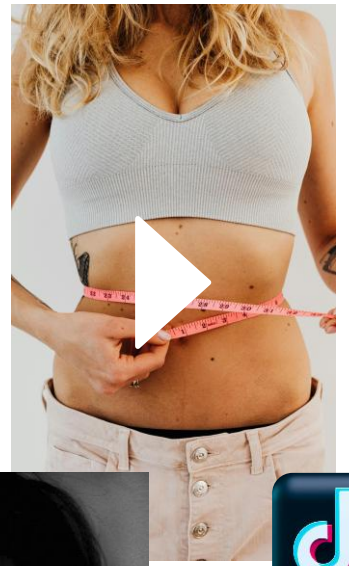
5. HOUR - harmful content **65%** out of 100%

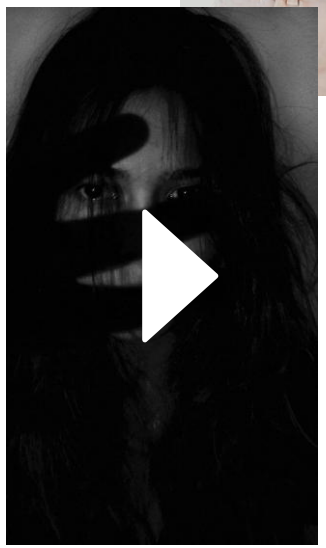
- **Even more sad videos appear on Ella's screen.**
- Videos that glorify eating disorders and now directly give concrete advice on how to limit eating.



THE END OF BROWSING - harmful content **95%** out of 100%

- At the end of the browsing process, the Tiktok algorithm had shown Ella a huge number of videos related to depression, suicidal tendencies, low-calorie diets and eating disorders.





Yle, 11 de noviembre de 2023

FUENTE DE LAS IMÁGENES | Unsplash y Pexels

Sesgos algorítmicos y de aprendizaje automático

CASO – El algoritmo dañino de TikTok

Los resultados fueron alarmantes

Según la psicóloga social, Suvi Uski, el algoritmo muestra al usuario el contenido que le interesa tanto como sea posible, sin importar cuán dañino sea. Uski dice que esto es particularmente problemático porque, al mismo tiempo, Tiktok es, con mucho, la más adictiva de las aplicaciones sociales.

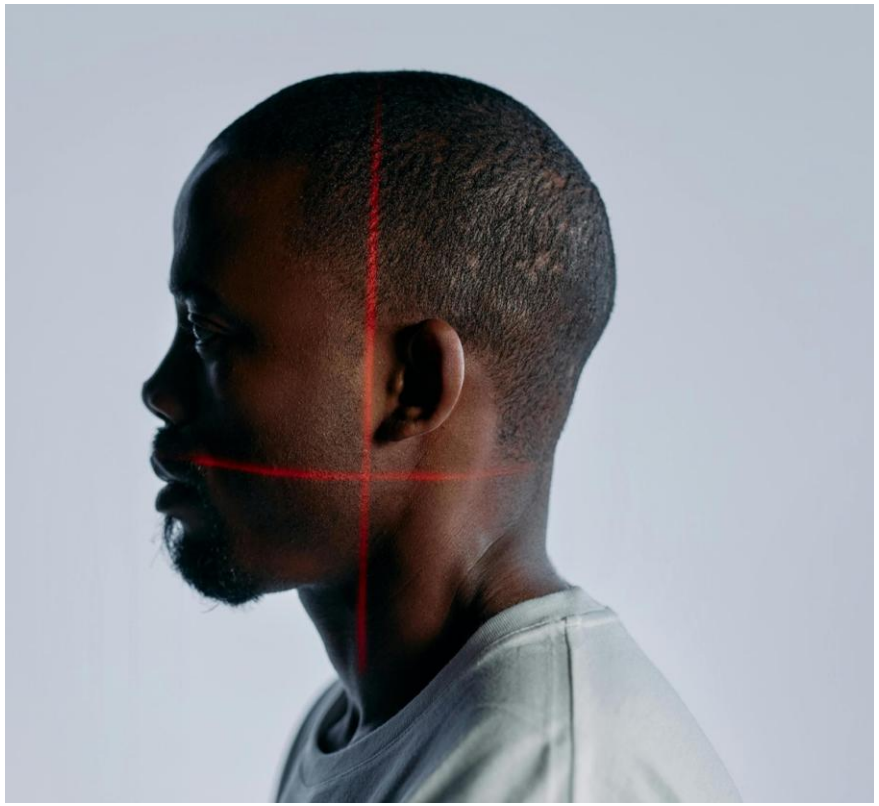
Yle, 11 de noviembre de 2023

Sesgos algorítmicos y de aprendizaje automático

Preocupaciones éticas relacionadas con el reconocimiento facial

Reconocimiento facial: ¿qué y cómo?

- Un sistema de reconocimiento facial escanea e identifica el rostro de una persona comparándolo con una imagen digital almacenada o un fotograma de un videoclip dentro de una base de datos utilizando un algoritmo de IA.
- La tecnología funciona identificando los rasgos faciales de una imagen y comparándolos con otras imágenes o vídeos de la base de datos mediante un algoritmo de IA.



FUENTE DE LA IMAGEN | Pexels

Shehmir Javaid, 12 de enero de 2024)

Sesgos algorítmicos y de aprendizaje automático

Reconocimiento facial: ejemplos de su uso

FUENTE DE LAS IMÁGENES | Unsplash y Pexels

1. **Atención sanitaria.** Escaneo automático de la cara del paciente mediante el reconocimiento facial para extraer el historial médico y la información del seguro. También se utiliza en el diagnóstico de trastornos médicos.

2. **Venta al por menor.** Amazon utiliza el reconocimiento facial para los pagos sin contacto. Se puede utilizar para personalizar la experiencia de compra, por ejemplo, para ofertas de fidelización de clientes. También se puede utilizar para mejorar la seguridad en el comercio minorista.

3. **Banca y finanzas.** Se utiliza en la identificación y verificación de clientes. Para acelerar el proceso de incorporación de clientes. Y para reducir los robos o fraudes bancarios.

4. **Servicios policiales y de seguridad.** En la identificación y registro de delincuentes y personas de interés. Para detectar fugitivos mediante el uso de cámaras de vigilancia en una ciudad, por ejemplo, en cruces, estaciones de tren y autobús y aeropuertos..

Sesgos algorítmicos y de aprendizaje automático

CASE – Recopilación de datos de reconocimiento facial poco ética

Vídeo sobre reconocimiento facial en aeropuertos en el canal de Youtube del Washington Post

Mira el video sobre el reconocimiento facial en los aeropuertos.

¿Qué cuestiones éticas surgen durante el video?

Washington Post, 10 de junio de 2019

Sesgos algorítmicos y de aprendizaje automático

Preocupaciones éticas sobre la recopilación de datos de reconocimiento facial

"El FBI y el ICE descubren que las fotos de las licencias de conducir del estado son una mina de oro para las búsquedas de reconocimiento facial"

Lea el artículo de noticias en The Washington Post el 7 de julio de 2019

Agentes de la Oficina Federal de Investigaciones (FBI, por sus siglas en inglés) y del Servicio de Inmigración y Control de Aduanas (ICE, por sus siglas en inglés) han utilizado las bases de datos de licencias de conducir del estado como una rica fuente de reconocimiento facial y han escaneado las fotos de millones de estadounidenses sin su conocimiento o consentimiento.

Drew Harwell, The Washington Post, 7 de julio de 2019

Sesgos algorítmicos y de aprendizaje automático

Sesgo racial en los sistemas de reconocimiento facial

"Las personas asiáticas y afroamericanas tenían hasta 100 veces más probabilidades de ser identificadas erróneamente que los hombres blancos, dependiendo del algoritmo particular y el tipo de búsqueda".

Lea la noticia en The Washington Post el 19 de diciembre de 2019

Un estudio federal confirmó que muchos de los sistemas de reconocimiento facial tienen prejuicios raciales.

Drew Harwell, The Washington Post, 19 de diciembre de 2019

HERRAMIENTAS PARA DETECTAR SESGOS ALGORÍTMICOS

FUENTE DE LA IMAGEN | Freepik

Herramientas para detectar sesgos algorítmicos

Evaluación del impacto ético de la UNESCO

- Desarrollado por la UNESCO, 2023
- Una herramienta para llevar a cabo una evaluación ética mediante la evaluación e identificación de los beneficios y riesgos de los sistemas de IA
- La herramienta consta de diversas preguntas que están listas para ser completadas
- Preguntas abiertas y de opción múltiple
- Ética considerada desde el diseño a lo largo de la evaluación
- Comenzando con preguntas exploratorias seguidas de preguntas relacionadas con los principios de la UNESCO.

La herramienta te ayudará a responder preguntas como::

- ¿Quién interactuará con la solución de IA?
- ¿Cuáles son las funciones y responsabilidades del equipo de IA?
- ¿Quiénes son las partes interesadas que se verán afectadas por el sistema de IA?
- ¿Cómo se considera la seguridad?
- Cualquier problema relacionado con la sostenibilidad, la privacidad, la transparencia, la explicabilidad y la rendición de cuentas
- **Encuéntralo aquí**

Figure 2: **Positionality Matrix (developed by The Alan Turing Institute)**

UNESCO, 2023

FUENTE DE LA IMAGEN | UNESCO

Herramientas para detectar sesgos algorítmicos

Evaluación del impacto ético de la UNESCO

¿Cómo detecta el sesgo la herramienta?

- La sección Equidad, No discriminación, Diversidad se centra más en detalle en los sesgos que aparecen en los datos
- Los sesgos también se abordan prestando atención a las funciones y responsabilidades.

6.2. Procedural Assessment

Throughout this section, when responding to questions about testing with particular groups, the project team should consider especially – but not only – race, colour, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, and disability. Please specify if testing was conducted on groups that combine several of these criteria i.e. if the system has been tested in terms of intersectionality.

6.2.1. Preventing discriminatory outcomes:

6.2.1.1. Has the algorithm been tested with different groups?

6.2.1.1.1. Was there a difference in terms of accuracy rate (or any other performance metric used)? Please describe any difference of this kind.

6.2.1.1.2. Was there a discriminatory effect for particular groups?

6.2.2. Data quality and preventing discriminatory bias:

6.2.2.1. Are processes in place to test data against biases?

6.2.2.1.1. Have you undertaken an analysis of the data to prevent societal and historical biases in data?

6.2.2.1.2. Is the data well-balanced and does it reflect the diversity of the targeted end-user population?

6.2.2.1.3. Are there any differences you can foresee between the data used for training and the data processed by the AI system which could result in the AI system producing discriminatory outcomes or performing differentially for different groups?

6.2.2.1.4. Have you developed a process to document how data quality issues can be resolved during the design process?

6.2.2.1.5. Did you put in place educational and awareness initiatives to help AI designers and developers gain awareness of the possible bias they can introduce through the design and development of the AI system?

6.2.2.1. Are processes in place to test data against biases?

UNESCO, 2023

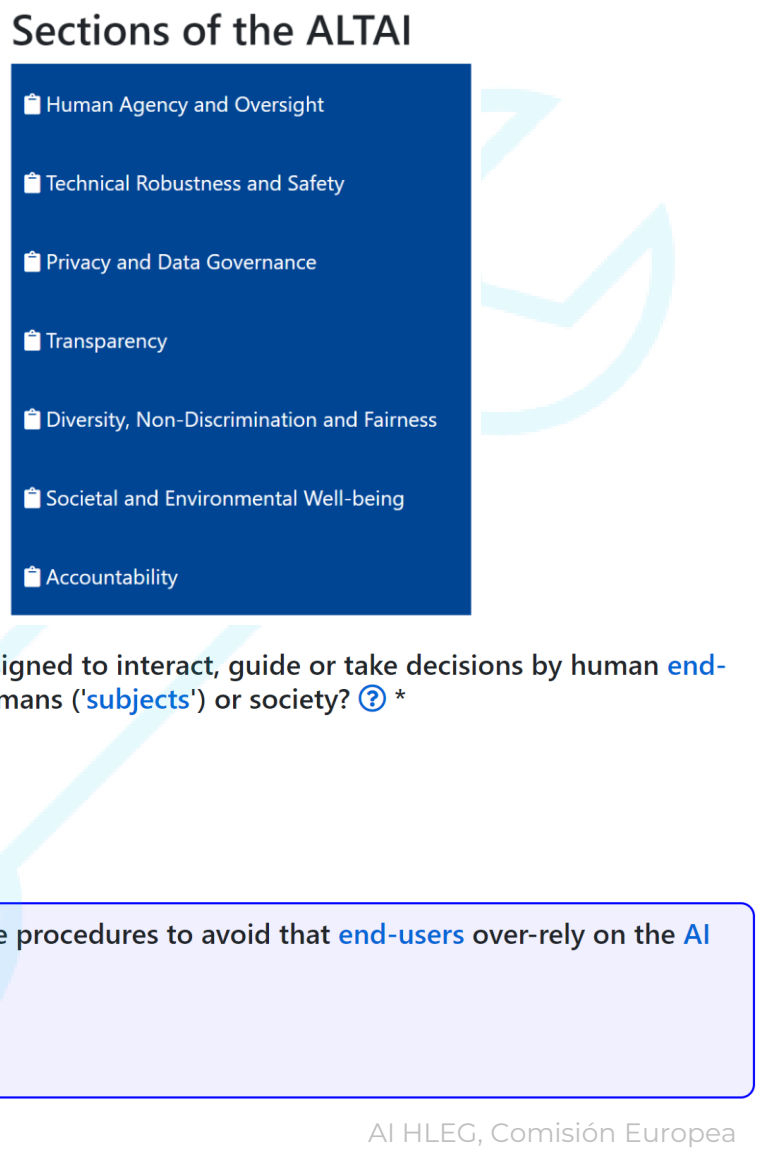
FUENTE DE LA IMAGEN | UNESCO

Herramientas para detectar sesgos algorítmicos

ALTAI

- Lista de evaluación de IA confiable
- Desarrollado por el Grupo de Expertos de Alto Nivel (AI HLEG) sobre inteligencia artificial, Comisión Europea
- La Lista de Evaluación para la IA Confiable (ALTAI)
- Lista de 7 requisitos para una IA confiable que se definen más específicamente en el sitio web
- Incluye una recomendación sobre cómo completar el ALTAI y quién debe participar

- Encuéntralo aquí



AI HLEG, Comisión Europea

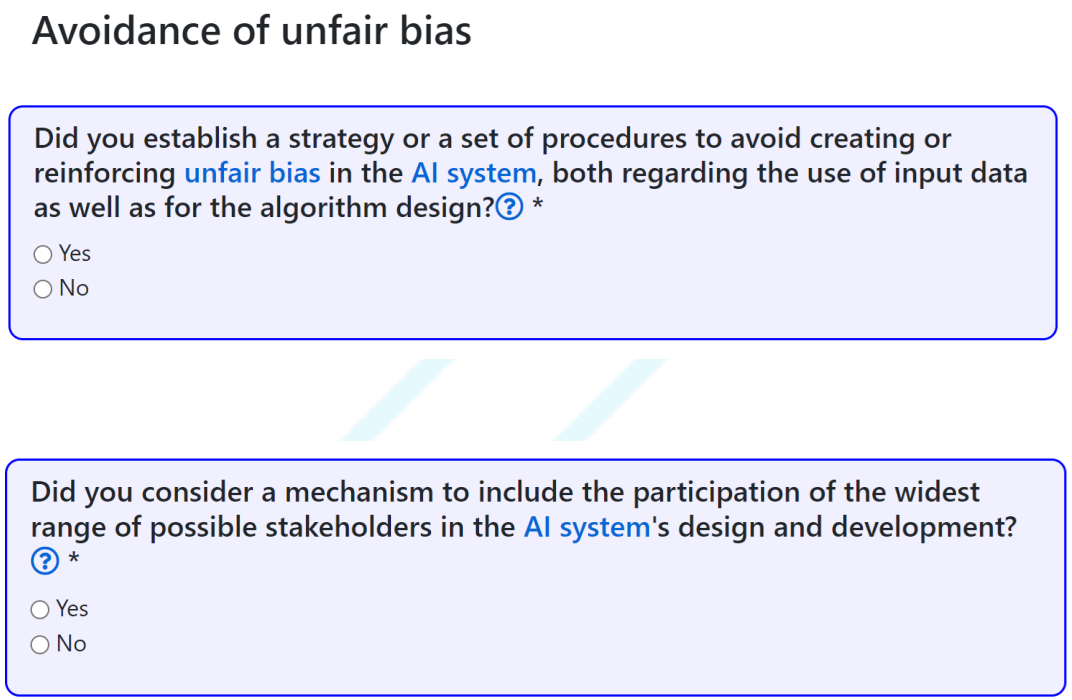
FUENTE DE LA IMAGEN | AI HLEG, Comisión Europea

Herramientas para detectar sesgos algorítmicos

ALTAI

¿Cómo detecta el sesgo la herramienta?

- La sección Diversidad, No discriminación y Equidad de la herramienta ALTAI se centra en evitar los sesgos injustos.
- La herramienta ALTAI evalúa el sesgo algorítmico a través de preguntas específicas, evaluando los datos de entrada y el diseño del algoritmo para detectar sesgos, al tiempo que promueve la inclusión, la accesibilidad y la participación de las partes interesadas para garantizar la equidad y la transparencia en los sistemas de IA.



AI HLEG, Comisión Europea

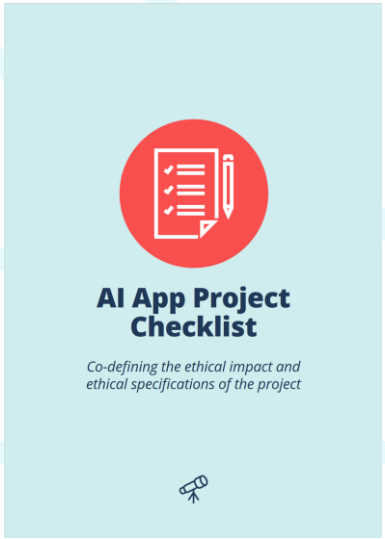
FUENTE DE LA IMAGEN | AI HLEG, Comisión Europea

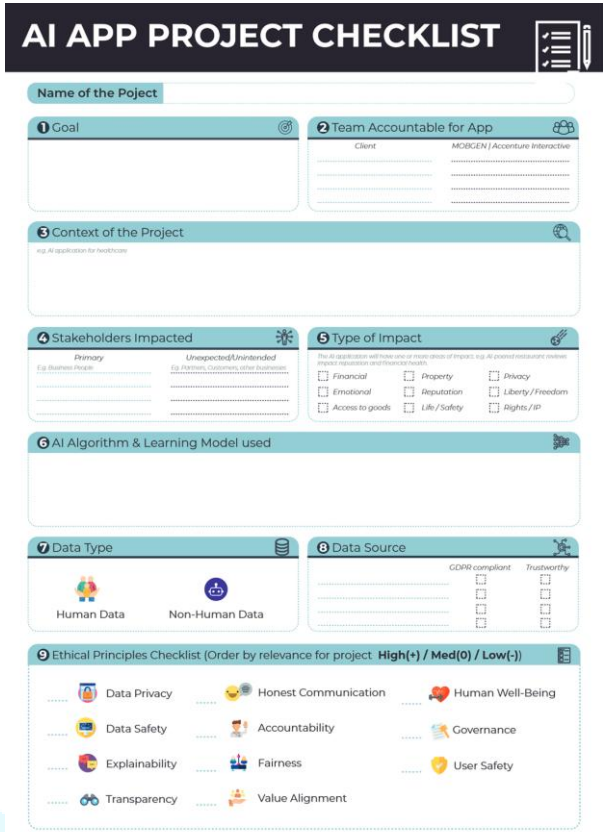
Herramientas para detectar sesgos algorítmicos

Kit de herramientas éticas para el desarrollo de aplicaciones de IA

- Desarrollado por Mario Sosa Hidalgo, (TU Delft Industrial Design Engineering)
- Tiene en cuenta la ética en el desarrollo de aplicaciones de IA
- Kit de herramientas innovador que tiene un enfoque activo, concreto y visual

[Encuéntralo aquí]





Mario Sosa Hidalgo, 2019

FUENTE DE LA IMAGEN | Mario Sosa Hidalgo, 2019

Herramientas para detectar sesgos algorítmicos

Kit de herramientas éticas para el desarrollo de aplicaciones de IA

¿Cómo detecta el sesgo la herramienta?

La herramienta de Mario Sosa aborda los sesgos y recomienda a los desarrolladores que se aseguren de que el sistema de IA desarrollado sea lo más imparcial posible, recordando las estrategias para los fallos de equidad.

FUENTE DE LA IMAGEN | Mario Sosa Hidalgo, 2019

PROYECTO | LA IA EN EL RECLUTAMIENTO

FUENTE DE LA IMAGEN | Freepik

Proyecto | La IA en el reclutamiento

?

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO

Estudio de caso | Fondo

Una empresa industrial global se enfrenta a ineficiencias e incertidumbres en su proceso de contratación, lo que supone una presión significativa sobre el departamento de RRHH al consumir tiempo y recursos. Anticipando el crecimiento futuro, la compañía también reconoce la necesidad urgente de contratar una variedad de especialistas, incluidos ingenieros, comercializadores, profesionales de finanzas, codificadores y gerentes de proyectos. La feroz competencia por empleados competentes y los altos costos asociados con los errores de reclutamiento enfatizan la importancia de un reclutamiento exitoso. La situación aumenta la presión sobre el equipo de RRHH, cuya tarea es garantizar el éxito y el buen funcionamiento del proceso de contratación. Como resultado, la compañía está explorando el desarrollo de una solución de inteligencia artificial adaptada a sus necesidades de contratación. Al integrar la IA en el proceso de contratación, el objetivo principal es mejorar la eficiencia, la objetividad y la tasa de éxito general de la contratación. Como primer paso, se consideran los pros y los contras de una solución de contratación de IA.

Proyecto | La IA en el reclutamiento

Contratación de IA | PROS y CONTRAS

PROS

•Eficiencia y gestión del tiempo: la IA agiliza las tareas, centra a los reclutadores en la estrategia y acelera la contratación

•Mejora de la calidad de los candidatos: la IA analiza los datos para seleccionar a los candidatos más adecuados, lo que reduce la rotación de personal

•Mejora de la experiencia del solicitante: la IA proporciona actualizaciones oportunas, lo que mejora la comunicación

•Reducción de costes: La IA reduce los costes de contratación al reducir el trabajo manual en las tareas rutinarias

•Mayor equidad: La IA reduce los sesgos, mejorando la objetividad en la contratación > mejora de la diversidad

•Análisis perspicaz de las redes sociales: la IA evalúa las redes sociales para determinar si encajan culturalmente

•Reclutamiento global: la IA supera las barreras lingüísticas y accede a grupos de talento más amplios

CONTRAS

•Reducción del compromiso personal: la IA hace que la contratación sea más mecánica, lo que posiblemente disuada a los candidatos

•Riesgo de sesgo: El diseño inadecuado de la IA o los datos sesgados pueden comprometer la diversidad y la equidad, perpetuando las desigualdades sociales

•Seguridad de los datos: El tratamiento de datos personales por parte de la IA aumenta los riesgos de privacidad y ciberataques

•Altos costos de inversión: la IA en el reclutamiento exige importantes inversiones financieras y de recursos

•Juicio humano: la confianza en la IA podría perder candidatos cualificados debido a problemas de datos

•Cumplimiento legal: La IA de contratación debe adaptarse a las leyes cambiantes para mantener la equidad

•Evaluación cultural: Es posible que la IA no evalúe el ajuste cultural o el trabajo en equipo de manera efectiva

(www.recruiter.com, 2023; Bursell y Roumbanis, 2024; Reclutamiento Arivu, 2023; Köchling y Wehner, 2020, Albaroudi et al., 2024, Sheard, 2022; Albassam, 2023).

Proyecto | La IA en el reclutamiento

?

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO

Entendiendo el problema

La empresa comienza una investigación de fondo para desarrollar una solución de IA, formando un equipo interno dedicado a identificar los desafíos clave en el proceso de contratación existente. El equipo realiza entrevistas con una variedad de partes interesadas, incluido el departamento de recursos humanos, los empleados recién contratados y los gerentes que han incorporado recientemente nuevos miembros al equipo. El equipo también recopila datos sobre los costos actuales de contratación y analiza la tasa de éxito de estas iniciativas. Comprender las percepciones de los empleados potenciales sobre el uso de la IA en la contratación es crucial para la empresa. Su objetivo es evitar cualquier daño a su reputación o marca. Para lograr esto, la compañía se compromete a garantizar que la adopción de la inteligencia artificial no solo obtenga una aceptación generalizada, sino que también se adhiera a los estándares éticos y legales.

Proyecto | La IA en el reclutamiento

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO

Entendiendo a los usuarios del sistema

Una de las personas entrevistadas es la Lead Recruiter María, quien cuenta con varios años de experiencia en la empresa. Ha visto de primera mano la acumulación de tareas de selección y la creciente presión que provoca en el departamento de RRHH, que se intensifica a la hora de contratar talento global.

María es de mente abierta e interesada en las posibilidades de la IA para ayudar en las tareas de contratación, aunque no está muy familiarizada con la tecnología de inteligencia artificial.

HERRAMIENTA

Las personas son perfiles de usuario detallados en diseño que encarnan los rasgos clave de los grupos de usuarios objetivo, evitando los estereotipos. Incluyen comportamientos, necesidades y datos demográficos para fomentar la empatía y guiar a los equipos de diseño en la creación de soluciones centradas en el usuario. Son útiles para trascender los segmentos de mercado convencionales y obtener resultados de diseño más inclusivos. (Kumar, 2013; Stickdorn et al., 2018a, 2018b)

Proyecto | La IA en el reclutamiento

Maria – RECLUTADORA PRINCIPAL

#La motivacióndetermina
#Buenos talentos

FUENTE DE LA IMAGEN | Imagen realizada por VAMK

Empática

Mente abierta

Saliente

Jugadora de equipo

Perfeccionista

Datos Demográficos

Edad: 37

Ubicación: Helsinki, Finlandia

Nacionalidad: Guatemalteca

Formación: MBA en Gestión de Recursos Humanos

Ocupación: Reclutador principal

Familia: Vive sola con su gato

OBJETIVOS DE TRABAJO

En busca de nuevos talentos

Satisfacción de los empleados

Proceso de reclutamiento rápido

HOBBIES

Natación

Escalada

Lectura

FRUSTRACIONES

Alta carga de trabajo

Encontrar un buen talento, que se aleje justo antes de comenzar el trabajo

Proceso de reclutamiento demasiado lento que hace que los candidatos se vayan a otra parte.

No hay tiempo suficiente para mantener informados a los candidatos durante el proceso de contratación

Siente presión para encontrar a los candidatos perfectos

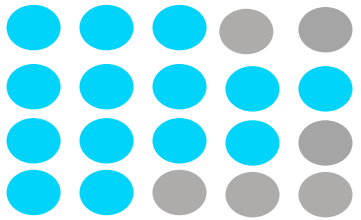
HABILIDADESSo

Medios de comunicación social

Software de RRHH

Entrevistando

Evaluación comparativa



María tiene grandes esperanzas, pero también le preocupa el impacto que tendrá en su trabajo la contratación impulsada por la IA. Le preocupa cómo la introducción de la IA puede cambiar las funciones del personal de RRHH en el futuro: ¿Qué nuevas capacidades y competencias necesita poseer? ¿Se valorará menos su contribución? ¿Sigue siendo necesaria su experiencia? ¿Y hasta qué punto se puede automatizar el proceso de reclutamiento?

Proyecto | La IA en el reclutamiento

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO

Comprender el proceso de contratación

Para obtener una comprensión integral del proceso de reclutamiento, la empresa desarrolla un viaje de reclutamiento que involucra a María y al equipo de reclutamiento. En colaboración con los gerentes de línea, navegan a través de cada etapa del viaje para identificar frustraciones significativas y pasos que requieren mucho tiempo. Además, exploran la posible integración de la IA en cada etapa para mejorar la eficiencia y la eficacia del proceso de contratación

HERRAMIENTA

Un mapa de viaje es una herramienta de diseño que describe la experiencia de servicio paso a paso de un usuario, incluidas las etapas de interacción, los puntos de contacto y los desafíos. También capta las respuestas emocionales para resaltar la calidad de la experiencia. Los mapas de recorrido pueden representar interacciones actuales o futuras y van desde amplias descripciones generales hasta instantáneas detalladas. (Kumar, 2013; Stickdorn et al., 2018a)

Proyecto | La IA en el reclutamiento

El viaje de Maria: reclutamiento

IDENTIFICACIÓN DE LAS NECESIDADES DE CONTRATACIÓN

La necesidad de una nueva persona es identificada por el gerente de línea y se informa a María

PREPARACIÓN DE LA DESCRIPCIÓN DEL PUESTO

La descripción del puesto se crea para aclarar las habilidades específicas necesarias

BÚSQUEDA DE TALENTO

María necesita identificar y atraer nuevos talentos

SELECCIÓN Y PRESELECCIÓN

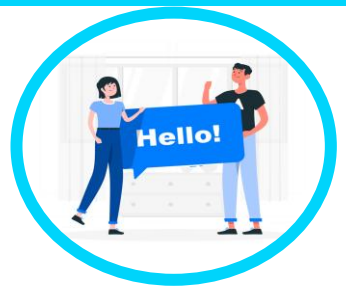
María necesita identificar a los candidatos adecuados de un gran grupo de candidatos

ENTREVISTAS

María y el superior jerárquico entrevistan a buenos candidatos

EVALUACIÓN Y OFERTA

Selección y verificación de antecedentes con referencia. La oferta se entrega al candidato.



INTRODUCCIÓN E INDUCCIÓN

María envía el kit de inducción a empleado y la inducción continúa cuando el empleado llega

Recruiter.com; Chen, 2022; Hunkenschroer y Luetge, 2022, Lawton; Albouradi 2024; Marr 2023; Fritts y Carbera, 2021; Vivek, 2023

FUENTE DE LA IMAGEN | Elaboracion propia

Proyecto | La IA en el reclutamiento

Cómo se puede utilizar la IA en diferentes etapas

IDENTIFICACIÓN DE LAS NECESIDADES DE CONTRATACIÓN

- Análisis de factores externos como las demandas del mercado, las acciones de la competencia y las métricas de RRHH
- Evaluar las tendencias laborales históricas, la demografía y los requisitos de habilidades junto con evaluaciones internas
- Seguimiento de métricas como las tasas de rotación y el tiempo de cobertura de los puestos para la eficiencia de la contratación

PREPARACIÓN DE LA DESCRIPCIÓN DEL PUESTO

- Mitigar el sesgo humano en la creación de descripciones de trabajo
- Ajustar el lenguaje en los anuncios y monitorear el impacto en el volumen y la calidad de la aplicación

BÚSQUEDA DE TALENTO

- Permitir una comunicación personalizada en plataformas digitales y redes sociales
- Bots de IA que identifican a los solicitantes de empleo activos y pasivos en todas las plataformas y bases de datos de solicitantes anteriores
- Emplear estrategias de minería de texto para mejorar el atractivo de la oferta de empleo con frases que atraigan a los candidatos.

SELECCIÓN Y PRESELECCIÓN

- Cribado inicial mediante el análisis de la semántica y la predicción de comportamientos potenciales y negativos a largo plazo
- Entrevistas en video mediante la evaluación de señales verbales y no verbales para evaluar la personalidad y las habilidades cognitivas
- Análisis de redes sociales mediante el examen de los comportamientos en línea para el ajuste cultural
- Análisis de currículum vitae con técnicas de PNL que enriquecen las evaluaciones de currículum

Recruiter.com; Chen, 2022; Hunkenschroer y Luetge, 2022, Lawton; Albouradi 2024; Marr 2023; Fritts y Carbera, 2021; Vivek, 2023

Proyecto | La IA en el reclutamiento

Cómo se puede utilizar la IA en diferentes etapas

ENTREVISTAS

- Técnicas de entrevista con IA utilizadas principalmente en la fase de preselección

EVALUACIÓN Y OFERTA

- Analizar los puestos de trabajo anteriores del candidato para crear una oferta personalizada que el candidato probablemente acepte.

INTRODUCCIÓN E INDUCCIÓN

- Automatización de la creación de documentos
- Seguimiento del progreso de la incorporación
- Adaptación de las trayectorias educativas

Recruiter.com; Chen, 2022; Hunkenschroer y Luetge, 2022, Lawton; Albouradi 2024; Marr 2023; Fritts y Carbera, 2021; Vivek, 2023

Proyecto | La IA en el reclutamiento

FASE DE DISEÑO

Reconocer los sesgos personales

Después de una declaración clara del problema, la empresa ha optado por proceder con una solución impulsada por IA. Ha sido nombrado desarrollador principal de IA responsable de supervisar la implementación de esta tecnología de IA elegida. La compañía reconoce la importancia de reunir un grupo diverso y de múltiples partes interesadas para asegurar una variedad de perspectivas para el éxito del proyecto. En un esfuerzo por abordar los posibles sesgos inconscientes dentro del equipo, la compañía ha ordenado que todos los miembros participen en un ejercicio inspirado en Alan Turing llamado "Matriz de posicionalidad

HERRAMIENTA

La matriz de posicionalidad de Alan Turing está diseñada para iluminar los sesgos personales del equipo, demostrando cómo los antecedentes socioeconómicos, los contextos culturales y las experiencias de vida individuales pueden influir en el juicio y la toma de decisiones.

alan-turing-institute.github.io

Proyecto | La IA en el reclutamiento

Prueba de Alan Turing | Reflexiona sobre el propósito, la posición y el poder

Características Personales & Identificación de grupo

¿Cómo me identifico?
Edad, raza y etnia, estado de discapacidad, religión, género, sexualidad, estado civil, estado parental, antecedentes lingüísticos

Educación, Formación y Trabajo

¿Cómo me han educado y formado?
Escuelas a las que asistió, nivel de educación, oportunidades de avance y desarrollo profesional, historial de empleo

MATRIZ DE POSICIONALIDAD

¿Hasta qué punto mis características personales, identificaciones de grupo, estatus socioeconómico, formación y antecedentes laborales, composición del equipo y marco institucional representan fuentes de poder y ventajas de marginación y desventaja? ¿Cómo influye esta posición en mi capacidad (y en la de mi equipo) para identificar y comprender a las partes interesadas afectadas y los impactos potenciales de mi proyecto?

Marco Institucional y Composición del Equipo

¿Cómo es mi contexto institucional y la composición de mi equipo?
Estructura de autoridad dentro de mi equipo de proyecto, jerarquías de poder y propiedad de políticas más amplias en mi organización, niveles de autonomía en la toma de decisiones, oportunidades para expresar preocupaciones y objeciones, diversidad en el equipo, cultura de inclusión o exclusión.

Nivel socioeconómico

¿Cuál es mi historial socioeconómico?
Estatus socioeconómico en crecimiento, movilidad social a lo largo del tiempo, estado actual, aspiraciones socioeconómicas

Modificado de: Reflexionar sobre el propósito, la posición y el poder - Turing Commons (alan-turing-institute.github.io)

Proyecto | La IA en el reclutamiento

FASE DE DISEÑO

Mapa de partes interesadas

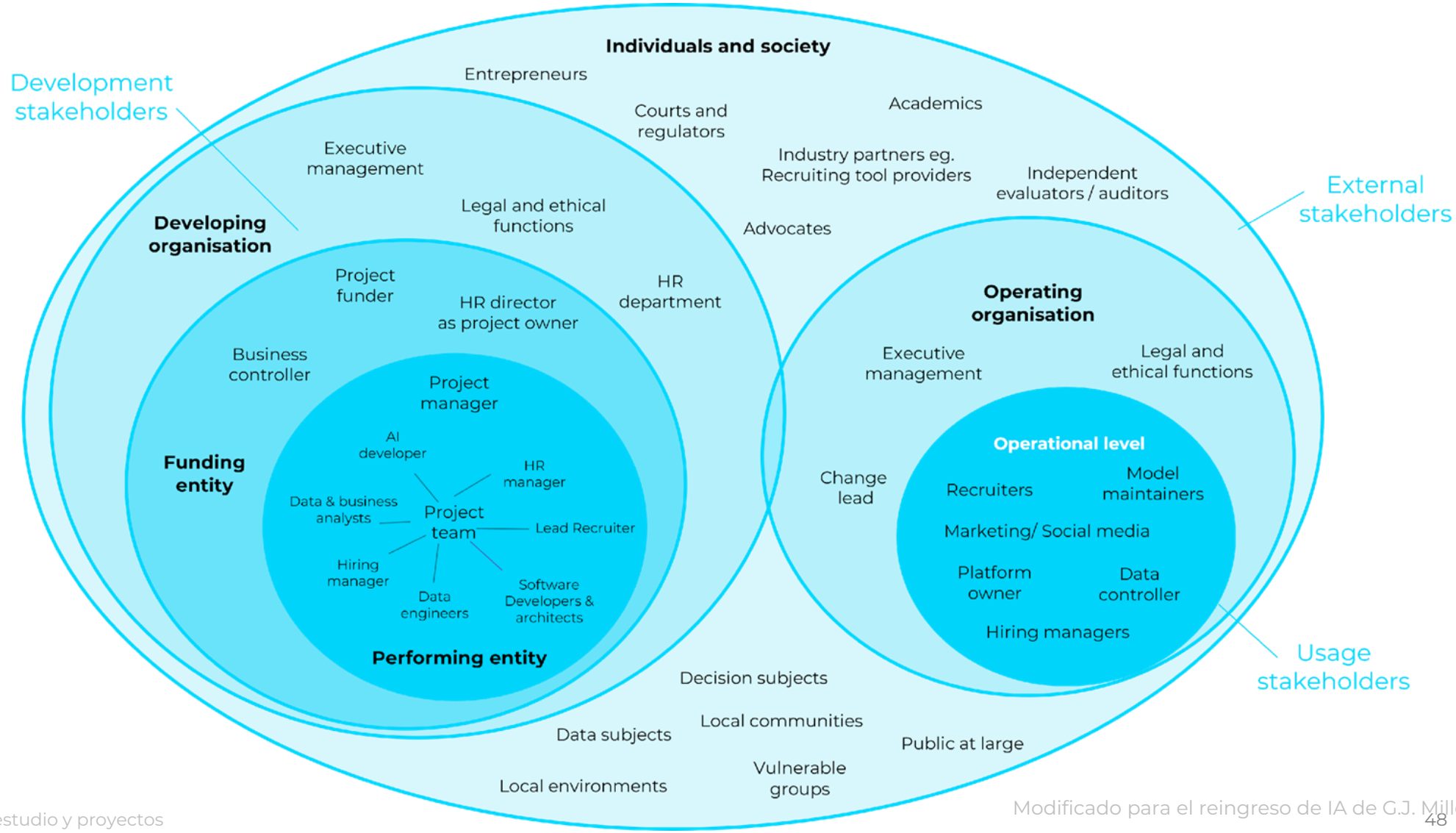
Además del grupo de partes interesadas internas que participan en el proyecto, la empresa también debe tener en cuenta a otras partes interesadas relacionadas. Por lo tanto, la empresa creó un mapa de partes interesadas siguiendo la definición de G.J. Miller del mapeo de partes interesadas del proceso de desarrollo de IA que define a las partes interesadas en el desarrollo y el uso, así como a las partes interesadas externas.

HERRAMIENTA

Un mapa de partes interesadas es una herramienta visual que se utiliza para identificar y mostrar las funciones y relaciones de todas las partes interesadas involucradas en un proyecto. Puede abarcar desde un simple cuadrante que muestra los niveles de influencia y compromiso hasta una matriz detallada que detalla las interacciones entre las partes interesadas. El mapa de partes interesadas ayuda a clarificar la dinámica de las partes interesadas y es crucial para comprender y gestionar las relaciones de la red. (Giordano et al., 2018; SDT, s.f.-b)

Proyecto | La IA en el reclutamiento

Mapa de partes interesadas Proyecto de reclutamiento de IA



Modificado para el reingreso de IA de G.J. Miller, 2022

Proyecto | La IA en el reclutamiento

FASE DE DISEÑO

Matriz de análisis de grupos de interés

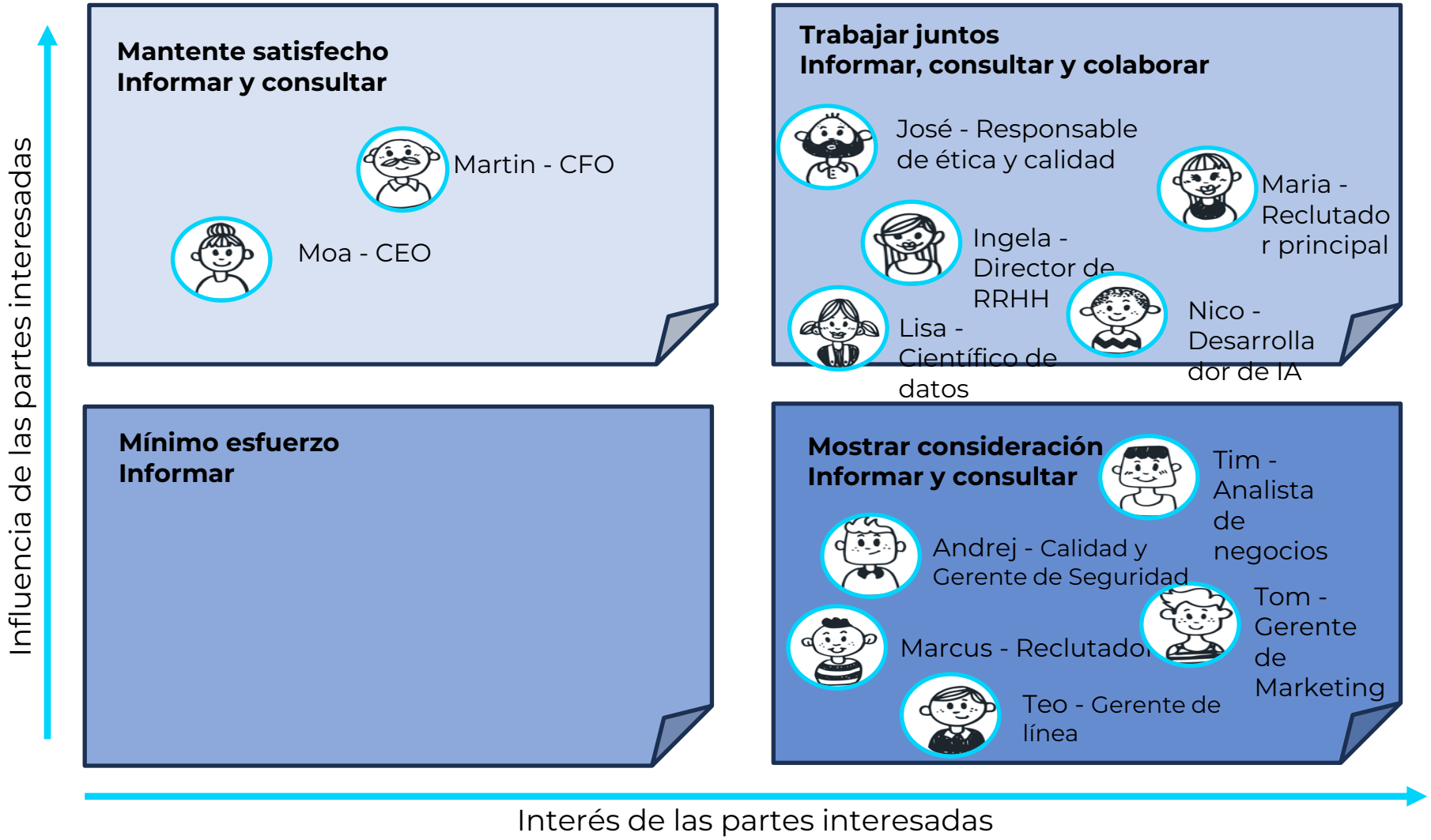
Sobre la base del mapa, el director del proyecto debe organizar a las partes interesadas en grupos para saber cómo mantenerlas informadas y ver cuál es su influencia en el proyecto.

HERRAMIENTA

La matriz de análisis de las partes interesadas organiza a las partes interesadas del proyecto por influencia e interés en una cuadrícula de cuadrantes para informar las tácticas de participación. Orienta cómo satisfacer a las partes interesadas de alta influencia y bajo interés; gestionar de cerca a los que tienen un gran interés e influencia; mantener informados a los interesados pero menos influyentes; y comprometer mínimamente a aquellos con poco de cualquiera de los dos. (Hoory y Botorff, 2022; Wallbridge, 2023)

Proyecto | La IA en el reclutamiento

Matriz de análisis de grupos de interés



Guía de estrategia de participación de las partes interesadas con ejemplos (kahootz.com)

FUENTE DE LAS IMÁGENES | Elaboración propia

Proyecto | La IA en el reclutamiento

FASE DE DISEÑO

Sesgos en las etapas de reclutamiento - Taller

Cuando los grupos de interés están claros, la empresa decide realizar un taller ético en el que se repasen todos los posibles sesgos en las diferentes etapas de la contratación. Se invita a las partes interesadas del campo "Trabajar juntos" (diapositiva anterior) a unirse. ¿Qué tipo de post-its habrías rellenado?

HERRAMIENTA

Un taller es una herramienta de diseño colaborativo en la que un grupo de participantes realiza actividades interactivas para abordar un problema específico o trabajar en un proyecto. Por lo general, facilitados por un líder, los talleres pueden variar en duración desde unas pocas horas hasta varios días, dependiendo de los objetivos y la complejidad de las tareas en cuestión. Este método es eficaz para generar ideas, resolver problemas y fomentar el trabajo en equipo y la innovación dentro de un grupo. (Wirtz, 2022)

Proyecto | La IA en el reclutamiento

Sesgos en la identificación de las necesidades de contratación

SESGO DE ANCLAJE
La persona cuyo puesto se va a cubrir afecta a las expectativas y a la predicción de las necesidades de contratación

Predicción de > de bajo rendimiento: necesidad de más recursos
Predicción de > de alto rendimiento: necesidad de menos recursos

SESGO DEL STATU QUO
Es posible que haya resistencia a cambiar el tipo de roles en la organización, incluso si las necesidades comerciales pueden indicar lo contrario

SESGO DEMOGRÁFICO
Es posible que los modelos predictivos no tengan en cuenta los cambios demográficos, como el envejecimiento de la población

SESGO DE BUCLE DE RETROALIMENTACIÓN
El algoritmo de IA está aprendiendo de sus propias decisiones pasadas, es decir. Si los datos de las necesidades de contratación están sesgados, podrían terminar creando un sesgo cada vez mayor, según sus primeras predicciones.

SESGO DE LA INDUSTRIA
En el caso de una visión demasiado estrecha de la industria, es posible que se pase por alto la necesidad futura más amplia.

Chen,2022; Vivek, 2023

Proyecto | La IA en el reclutamiento

Sesgos en la preparación de la descripción del puesto

Las descripciones de trabajo creadas por IA se generan principalmente para reducir el sesgo humano en la contratación, > la IA puede identificar el lenguaje que conduce a sesgos

SESGO HISTÓRICO
La descripción del trabajo se puede entrenar en función de descripciones de trabajo antiguas, que pueden estar sesgadas

SESGO DEMOGRÁFICO
La IA puede sugerir términos como "ambicioso" o "líder seguro", dirigiendo las descripciones a ciertos grupos demográficos.

SESGO DE ANCLAJE
En caso de que las descripciones de los puestos de trabajo se creen en función de las competencias básicas de los empleados de alto rendimiento, las descripciones de los puestos de trabajo pueden crear sesgos

Ongig.com; Lawton, 2022; Chen, 2023

Proyecto | La IA en el reclutamiento

Sesgos en la búsqueda de talento

Sesgo de GÉNERO, RACIAL y EDAD
Criterios de preselección para la adición seleccionada por el empleador excluyendo ciertos datos demográficos.

La rentabilidad en la entrega de anuncios puede llevar a mostrar los anuncios de empleo a los que probablemente hacen clic en el añadir en lugar de a los que tendrían éxito en el puesto.

La selección de la plataforma elegida por parte del empleador puede estar sesgada, ya que algunas plataformas pueden ser más utilizadas por ciertos grupos de solicitantes.

El uso de las redes sociales difiere mucho entre las personas; mientras que algunos pueden no tener tiempo para usar, por ejemplo. En LinkedIn, algunos pueden dar una imagen falsa de sí mismos, mientras que otros tienen más posibilidades de impactar en su huella digital.

El comportamiento del usuario está reforzando los algoritmos para replicar los patrones, es decir. Si cierto tipo de hombre blanco abre los anuncios de trabajo con más frecuencia, el anuncio será visible para usuarios similares con más frecuencia.

Si los reclutadores se comunican más con los candidatos de ciertas plataformas, los algoritmos pueden aumentar la visibilidad de los anuncios en esas plataformas, aumentando aún más el sesgo.

Sheard, 2022; Naakka 2018; Lawton, 2022; Chen 2022; Köchling y Wehner, 2022

Proyecto | La IA en el reclutamiento

Sesgos en el cribado y la preselección

En general, la IA puede generar un entorno de contratación más justo al reducir el sesgo inconsciente, que los humanos tienden a tener en función de la demografía.

Si los datos de capacitación existen en función de un grupo de empleados existentes o de un grupo estrictamente definido, pueden dar lugar a una discriminación involuntaria contra los grupos infrarrepresentados.

El sesgo de agregación puede ocurrir si se hacen generalizaciones falsas sobre poblaciones enteras de personas; Por ejemplo, "los jóvenes tienen habilidades tecnológicas", > deja fuera a las personas mayores.

Utilizar la huella digital del candidato como parte de la selección; Cuando el algoritmo se basa en preferencias como "energético" o "activo", puede discriminar a las personas discapacitadas.

Albouradi et al., 2024; Hunkenschoer y Luetge, 2022; Digitalocean.com; Vivek, 2023,

Proyecto | La IA en el reclutamiento

Sesgos en las entrevistas

Las técnicas de IA en las entrevistas se utilizan principalmente en la fase de preselección.

Las entrevistas finales deben realizarse en interacción de persona a persona.

En las evaluaciones de video, el software de reconocimiento emocional puede interpretar, por ejemplo. entonaciones en diferentes idiomas de manera diferente.

Es posible que algunos candidatos no se sientan seguros frente a la cámara.

¿Deshumanización? Socavar la relación entre el empleado y el empleador

El uso de la tecnología de reconocimiento facial puede generar preocupaciones sobre la privacidad y sesgos contra ciertos grupos*

*** Ver el ejemplo de los sesgos de reconocimiento facial**

Sheard, 2022; Naakka 2018; Lawton, 2022; Chen 2022; Köchling y Wehner, 2022

Proyecto | La IA en el reclutamiento

Creación e incorporación de ofertas

A tener en cuenta; ¿En qué etapa es necesaria la interacción humano-humano? El onboarding es dar una imagen de la empresa.

Incorporación
Los datos de entrenamiento pueden basarse en datos históricos y, por lo tanto, conllevar sesgos.

Incorporación
Los sesgos personales del creador del algoritmo de IA pueden afectar a la oferta educativa.

Las herramientas que ofrecen a menudo utilizan datos históricos para anticipar los estándares de la industria, > pueden generar sesgos.

Proyecto | La IA en el reclutamiento

RECOPILACIÓN Y PREPARACIÓN DE DATOS

Selección y preselección de los candidatos; Definición de conjunto de datos

Si bien la empresa está estudiando todos los aspectos y posibilidades en la contratación de IA, actualmente la empresa está evaluando intensamente la IA en la fase de selección y preselección. Esto implica un análisis en profundidad del escaneo de currículums y la alineación de los candidatos con las vacantes de trabajo adecuadas. Mediante la implementación de herramientas de IA para automatizar el proceso de selección de currículums, el objetivo es identificar de forma rápida y precisa a los candidatos que cumplen los requisitos del puesto. Se espera que la adopción de dicha tecnología conduzca a un uso más estratégico de los recursos de recursos humanos, mejore el calibre de los candidatos que llegan a la lista de finalistas y, al alinear las habilidades y antecedentes de los candidatos con los requisitos de la empresa, aumente la efectividad general del proceso de contratación. En primer lugar, la empresa debe definir los conjuntos de datos, que son la base del sistema de contratación y de los que los algoritmos toman los datos.

Proyecto | La IA en el reclutamiento

Entrenamiento de IA

- Los conjuntos de datos de entrenamiento desempeñan un papel crucial en el desarrollo de algoritmos. Estos conjuntos de datos son la base sobre la que se construyen los algoritmos. El propósito de estos conjuntos de datos es servir como la "verdad fundamental" o base fáctica que instruye al algoritmo sobre cómo interpretar y procesar grandes cantidades de datos.
- Enseñan al algoritmo a comprender las relaciones entre las variables de entrada (la información introducida en el sistema) y una variable de salida (la predicción o decisión que toma el sistema en función de las entradas).
- Un desafío significativo surge cuando los propios datos de entrenamiento contienen sesgos. Estos sesgos pueden ser el reflejo de desigualdades históricas, prejuicios o simplemente el resultado de una recopilación de datos no representativa o incompleta.
- Cuando un algoritmo se entrena con datos sesgados, aprende estos sesgos y los perpetúa en sus predicciones y decisiones. Este fenómeno se describe comúnmente como el problema del "sesgo hacia adentro, sesgo hacia afuera". Esto significa que si los datos de entrada (sesgo de entrada) están sesgados, la salida del modelo de aprendizaje automático (sesgo de salida) también estará sesgada.
- Esto puede dar lugar a resultados injustos, discriminatorios o defectuosos que afecten a los candidatos que se seleccionan o rechazan en función de criterios sesgados en lugar de sus verdaderas calificaciones o potencial.

Sheard, 2022

Proyecto | La IA en el reclutamiento

Entrenar a la IA con los datos internos y externos de la empresa

Datos internos

- **Datos de rendimiento de los empleados y datos históricos de empleo**
  - Datos recopilados internamente sobre el rendimiento de los empleados y las decisiones de contratación anteriores.
- **CV enviados a lo largo del tiempo y CV y currículos**
  - Currículum vitae histórico enviado directamente a la empresa para las solicitudes de empleo.
- **Datos de casos falsos negativos**
  - Información sobre candidatos rechazados indebidamente, derivada de los procesos de selección de la empresa.
- **Palabras clave y frases de CVs**
  - Palabras o frases específicas de los CV recopilados a través del proceso de solicitud de la empresa.
- **Información de Antecedentes Educativos**
  - Datos educativos obtenidos de las solicitudes de los candidatos a la empresa.

Big Data externo

- **Redes Sociales**
  - Plataformas en línea donde las personas comparten intereses personales, experiencias y logros profesionales. Las redes sociales pueden proporcionar información sobre la personalidad, las habilidades y la red profesional de un candidato, lo que las convierte en una fuente valiosa para identificar talentos potenciales.
- **Sitios web profesionales**
  - Sitios web diseñados específicamente para la creación de redes profesionales y el desarrollo profesional, como LinkedIn. Estos sitios contienen perfiles detallados que incluyen historial laboral, educación, habilidades, avales y logros profesionales, lo que ofrece una rica fuente de información para la identificación de talentos.
- **Búsquedas en Internet**
  - Búsquedas amplias en línea que las empresas pueden realizar para encontrar información disponible públicamente sobre un candidato. Esto podría incluir artículos de noticias, blogs personales, publicaciones o cualquier presencia en la web que proporcione contexto adicional sobre las calificaciones, los logros y el carácter de una persona.

Sheard, 2022; Fernandes, 2021; Ma, 2024

Proyecto | La IA en el reclutamiento

DESARROLLO Y FORMACIÓN DE MODELOS

Anuncio de empleo; Selección de palabras clave para el algoritmo

La empresa ha publicado recientemente una descripción del puesto de trabajo y ha recibido cientos de solicitudes en respuesta. El prototipo de selección y preselección de IA escaneará los CV de los solicitantes en busca de palabras clave y otra información que se considere relevante para las contrataciones exitosas, como experiencia, títulos laborales, empleadores anteriores, universidades y títulos.Sobre la base de esto, el sistema crea un perfil de solicitante estructurado y todos los candidatos son calificados y clasificados.

Anuncio de empleo y palabras clave relacionadas

We are looking for..

...Communications assistant to be responsible for the creation of content such as media releases, blogs, and social media posts on behalf of our company. You will also be monitoring media and campaign coverage and attending internal and external events.

RESPONSIBILITIES

- Develop and manage diverse content, including media releases, blogs, and social media posts, aligned with the company's strategic goals; monitor media and campaign coverage.
- Support the implementation of internal and external communications strategies and assist in managing the company's image.
- Organize marketing events and provide comprehensive administrative support, including maintaining event calendars and updating contact lists.
- Compile and prepare presentations and reports while tracking project progress and media exposure.

REQUIREMENTS

- Holds a Bachelor's degree in communications, marketing, or related field, with excellent verbal and written communication abilities.
- Proficient in social media strategies and media relations, with a creative and innovative approach.
- Strong organizational skills and attention to detail, capable of multitasking and maintaining positive interpersonal relationships.
- Skilled in using office management and design software, such as Photoshop and InDesign, and knowledgeable in various social media platforms.

"Licenciatura en comunicación"
"Licenciatura en marketing"
"Creación de contenidos"
"Comunicados de prensa"
"Blogs"
"Publicaciones en redes sociales"
"Monitoreo de medios" y "cobertura de campañas"
"Apoyo a la implementación de estrategias de comunicación"
"Gestión de la imagen de la empresa"
"Organización de eventos de marketing"
"Brindar apoyo administrativo integral"
"Mantenimiento de calendarios de eventos"
"Actualización de listas de contactos"
"Recopilación de presentaciones" e "informes"
"Seguimiento del progreso del proyecto" y "exposición en los medios"
"Competente en estrategias de redes sociales" y "relaciones con los medios"
"Enfoque creativo e innovador"
"Fuerte capacidad de organización" y "atención al detalle"
"Capaz de realizar múltiples tareas"
"Mantener relaciones interpersonales positivas"
"Experto en el uso de la gestión de oficinas" y "software de diseño"
"Photoshop" e "InDesign"
"Conocedor de varias plataformas de redes sociales"

FUENTE DE LA IMAGEN | Imagen desarrollada por VAMK

Proyecto | La IA en el reclutamiento

DESARROLLO Y FORMACIÓN DE MODELOS

Comparar los currículos de dos candidatos
Hacia la frase clave

A continuación, nos centraremos en analizar dos aplicaciones para entender cómo el prototipo de aplicación de IA evalúa a los candidatos durante la fase de selección. Este análisis ayudará a identificar si hay algún sesgo que influya en el proceso de selección. La empresa organiza un taller para explorar los detalles de la detección de CV impulsada por IA, utilizando un estudio de caso de sus propias operaciones. Durante el taller, analizan dos CV uno al lado del otro para descubrir cómo las diferencias en la redacción y la redacción pueden llevar inadvertidamente a pasar por alto a los candidatos calificados. Este ejercicio tiene como objetivo iluminar los posibles sesgos incorporados en sus herramientas de IA y fomenta el pensamiento crítico sobre el perfeccionamiento del sistema para reconocer y apreciar plenamente la amplitud de las capacidades de un solicitante.

CV 1

JOHN DOE

Communications assistant

About Me

A dedicated communications professional with a Bachelor's degree in Marketing, seeking the role of Communications Assistant to leverage extensive experience in content creation, social media management, and event coordination to contribute to the company's strategic goals.

+123-456-7890

hello@doeland.com

123 Anywhere St., Any City

EXPERIENCE

Communications assitant
ABC Corporation
2019- Present

- Developed and managed diverse content, including media releases, blogs, and social media posts
- Supported the implementation of comprehensive communications strategies
- Organized multiple marketing events and maintained event calendars
- Compiled presentations and tracked project progress, ensuring alignment with strategic goals

EDUCATION

University of communications
Bachelor's degree in Marketing
2014-2018

SKILLS SUMMARY

- Proficient in Photoshop and InDesign
- Skilled in social media strategies and media relations
- Excellent verbal and written communication abilities
- Strong organizational skills and attention to detail

CV 2

JANE DOE

Digital Content Creator

PROFILE

Passionate communicator with an academic background in Communications Studies and hands-on experience in crafting engaging narratives and managing digital platforms. Eager to bring my toolkit of creative dissemination and stakeholder engagement to the Communications Assistant position.

CONTACT ME

(123) 456-7890

Jane@doe.com

123 Anywhere St.,
Any city, State,
Country 12345

EDUCATION

CREATIVE UNIVERSITY
BSc in Communications Studies 2018

WORK EXPERIENCE

Digital Content Creator; Creative Media Co
2019 -Present

- Spearheaded content initiatives across digital channels, crafting engaging articles and dynamic social engagements
- Drove strategy for public relations efforts, enhancing brand visibility
- Led the logistics for promotional and networking gatherings, ensuring smooth execution
- Synthesized data into compelling reports and visuals for team updates

COMPETENCIES

- Advanced user of digital design tools and content management systems
- Crafted engaging online dialogue and cultivated media partnerships
- Articulate communicator, both in written formats and oral presentations
- Excelling in project orchestration and meticulous in administrative tasks

FUENTE DE LAS IMÁGENES | Imagen desarrollada por VAMK

Esta negrita ayuda a ver las conexiones entre la redacción específica de los CV y las herramientas de selección de IA basadas en palabras clave.

JOHN DOE

Communications assistant

About Me

A dedicated communications professional with a **Bachelor's degree in Marketing**, seeking the role of Communications Assistant to leverage extensive experience in **content creation, social media management**, and event coordination to contribute to the company's strategic goals.

+123-456-7890

hello@doeland.com

123 Anywhere St., Any City

EXPERIENCE

Communications assitant
ABC Corporation
2019- Present

- Developed and managed diverse content, including **media releases, blogs,** and **social media posts**
- Supported the implementation of comprehensive communications strategies
- Organized multiple **marketing events** and **maintained event calendar**s
- Compiled **presentations** and tracked **project progress,** ensuring alignment with strategic goals

EDUCATION

University of communications
Bachelor's degree in Marketing
2014-2018

SKILLS SUMMARY

- **Proficient in Photoshop** and **InDesign**
- Skilled in **social media strategies** and **media relations**
- Excellent verbal and written communication abilities
- **Strong organizational skills** and **attention to detail**

JANE DOE

Digital Content Creator

PROFILE

Passionate communicator with an academic background in Communications Studies and hands-on experience in crafting engaging narratives and managing digital platforms. Eager to bring my toolkit of creative dissemination and stakeholder engagement to the Communications Assistant position.

CONTACT ME

(123) 456-7890

Jane@doe.com

123 Anywhere St.,
Any city, State,
Country 12345

EDUCATION

CREATIVE UNIVERSITY
BSc in Communications Studies 2018

WORK EXPERIENCE

Digital Content Creator; Creative Media Co
2019 -Present

- Spearheaded **content** initiatives across digital channels, crafting engaging articles and dynamic social engagements
- Drove strategy for public relations efforts, enhancing brand visibility
- Led the logistics for promotional and networking gatherings, ensuring smooth execution
- Synthesized data into compelling **reports** and visuals for team updates

COMPETENCIES

- Advanced user of digital design tools and content management systems
- Crafted engaging online dialogue and cultivated media partnerships
- Articulate communicator, both in written formats and oral presentations
- Excelling in project orchestration and meticulous in administrative tasks

FUENTE DE LAS IMÁGENES | Imagen desarrollada por VAMK

Esta tabla ilustra el impacto de las diferencias de redacción y redacción en dos CV y cómo pueden afectar al resultado de un proceso de selección de IA. A pesar de poseer calificaciones relevantes, el segundo CV puede pasarse por alto debido a que sus descripciones no se alinean estrechamente con las palabras clave específicas establecidas por la empresa para la selección. Esto demuestra la importancia de hacer coincidir el idioma de un CV con las palabras clave esperadas por un sistema de IA.

ASPECTO	JOHN DOE	JANE DOE
Títulos de trabajo y terminología	Utiliza términos estándar como "Asistente de comunicaciones"	Utiliza títulos como "Creador de contenido digital"
Formación académica	Menciona explícitamente "Licenciatura en Marketing"	Dice "Licenciatura en Ciencias de la Comunicación"
Describir tareas y habilidades	Coincidencia directa con palabras clave como "Creación de contenido", "Comunicados de medios"	Utiliza un lenguaje variado como "elaborar narrativas atractivas", "iniciativas de contenido encabezadas"
Habilidades técnicas y de software	Especifica "Photoshop" e "InDesign""	Mención general de "herramientas de diseño digital y sistemas de gestión de contenidos"
Concordancias directas de palabras clave	Coincide estrechamente con muchas palabras clave especificadas	Utiliza diferentes frases que pueden no coincidir con las palabras clave específicas establecidas para la detección.

Proyecto | La IA en el reclutamiento

EVALUACIÓN Y PRUEBAS DE MODELOS

Probar la equidad de la solución

La evaluación del prototipo de algoritmo reveló inconsistencias en el análisis de CV, lo que provocó refinamientos para una comprensión más completa de los atributos de CV. El modelo ahora está preparado para una fase de prueba antes de la implementación en el mundo real. Si bien varios aspectos requieren evaluación, la atención se centrará aquí en probar la equidad de la aplicación para determinar el alcance de los sesgos incurridos durante el proceso. Las métricas de equidad se emplean para medir el nivel de sesgo presente en las decisiones de la aplicación.

HERRAMIENTA

El kit de herramientas de código abierto de IBM puede ayudar a examinar, informar y mitigar la discriminación y el sesgo en los modelos de machine learning a lo largo del ciclo de vida de las aplicaciones de IA. Revisa la demo y juega con la herramienta.

Equidad de IA 360 (ibm.com)

Proyecto | La IA en el reclutamiento

Probando la equidad

Para crear sistemas que sean equitativos para todos y libres de sesgos, como la discriminación racial o de género, es importante evaluar el rendimiento de la solución en diversos segmentos de la población.

Si bien existen varios métodos para evaluar la equidad de una solución, en este contexto de reclutamiento, nos centraremos exclusivamente en las pruebas de clasificación binaria.

Definir la equidad en términos matemáticos ha demostrado ser un desafío, lo que ha llevado a una falta de consenso sobre las formulaciones estándar. Sin embargo, la mayoría de las definiciones de equidad se unen en torno a varios conceptos clave:

- Desconocimiento
- Paridad demográfica
- Cuotas igualadas
- Paridad de tasa predictiva
- Equidad individual
- Equidad contrafáctica

La paridad demográfica, la paridad predictiva de tasas y las cuotas igualadas generalmente se agrupan bajo el paraguas de la "equidad de grupo". Este concepto se explorará más a fondo en esta sección, particularmente a través de la lente de la IA en el reclutamiento.

Wortman Vaughan y Wallace, 2022; Zhong, 2018

Proyecto | La IA en el reclutamiento

Matriz de confusión

Uno de los métodos de clasificación binaria es la matriz de confusión.

La matriz de confusión resume las predicciones realizadas por un modelo algorítmico en comparación con los resultados reales, en los que se entrenó. Muestra los recuentos de predicciones correctas e incorrectas, proporcionando una visión clara del rendimiento del modelo y su precisión explicable.

Matriz de confusión en el ejemplo de reclutamiento

	Incompetente	Cualificado
Rechazar	Verdadero negativo (TN) = candidatos no cualificados rechazados	Falso negativo (FN) = Candidatos rechazados, cualificados
Contratar	Falso positivo (FP) =Candidatos contratados, no cualificados	Verdadero Positivo (TP) =Candidatos contratados, cualificados

(Wortman Vaughan y Wallace, 2022; Zhong, 2018, Saplicki, 2022

Proyecto | La IA en el reclutamiento

Exactitud

En el ejemplo de reclutamiento, ilustramos la evaluación de la equidad utilizando el género como variable demográfica para la evaluación, asumiendo que hay 100 solicitantes mujeres y hombres.

HOMBRES	Incompetente	Cualificado
Rechazar	**15 (TN)**	5 (FN)
Contratar	20 (FP)	**60 (TP)**

MUJERES	Incompetente	Cualificada
Rechazar	**60 (TN)**	20 (FN)
Contratar	5 (FP)	**15 (TP)**

Exactitud (TP + TN)

(TP + FP + TN + FN)

Tanto en la categoría masculina como en la femenina, 75 aspirantes se clasifican correctamente
-> Es decir, la precisión es similar en ambos grupos demográficos

Proyecto | La IA en el reclutamiento

Paridad demográfica

Sobre la base de los datos de la matriz de confusión, la paridad demográfica es un criterio de equidad en el que la probabilidad de un resultado favorable (en nuestro ejemplo, la contratación) no se ve influida por una característica protegida, como el género.

HOMBRES	Incompetente	Cualificado
Rechazar	15 (TN)	5 (FN)
Contratar	**20 (FP)**	**60 (TP)**

MUJERES	Incompetente	Cualificada
Rechazar	60 (TN)	20 (FN)
Contratar	**5 (FP)**	**15 (TP)**

En el escenario de ejemplo de reclutamiento, si se contrata a 80 de cada 100 candidatos masculinos y solo a 20 de cada 100 candidatas mujeres, no se cumple la paridad demográfica. Sin embargo, esto podría ser aceptable si los solicitantes masculinos están realmente más calificados que las solicitantes femeninas, como parece ser en este caso.

Al mismo tiempo que se hace cumplir la equidad a nivel de grupo; Esto puede ser injusto para las personas, ya que podría abandonar a un candidato calificado solo para lograr la paridad demográfica

Proyecto | La IA en el reclutamiento

Paridad predictiva

A diferencia de la paridad demográfica, la paridad predictiva tiene en cuenta tanto las decisiones del clasificador como los resultados reales. Evalúa si la probabilidad de que un candidato sea adecuado para un puesto es coherente en los diferentes grupos, siempre que la IA lo haya seleccionado para su contratación. Se espera que esta probabilidad sea casi idéntica para todos los grupos considerados.

HOMBRES	Incompetente	Cualificado
Rechazar	15	5
Contratar	20	**60**

MUJERES	Incompetente	Cualificada
Rechazar	60	20
Contratar	5	**15**

En el ejemplo de reclutamiento, el sistema cumple con los criterios de paridad predictiva, ya que tres cuartas partes de los solicitantes contratados de ambos grupos están calificados.

Castelnovo et al., 2022; Wortman Vaughan y Wallace, 2022

Proyecto | La IA en el reclutamiento

Cuotas igualadas; Tasa de falsos positivos y tasa de falsos negativos

La métrica conocida como tasa de falsos positivos afirma que la probabilidad de que un candidato no calificado sea contratado erróneamente debe ser consistente en todos los grupos demográficos.

HOMBRES	Incompetente	Cualificado
Rechazar	15	5
Contratar	**20**	60

MUJERES	Incompetente	Cualificado
Rechazar	60	20
Contratar	**5**	15

La métrica conocida como tasa de falsos negativos afirma que la probabilidad de que un candidato calificado sea rechazado erróneamente debe ser consistente en todos los grupos demográficos.

HOMBRES	Incompetente	Cualificada
Rechazar	15	**5**
Contratar	20	60

MUJERES	Incompetente	Cualificada
Rechazar	60	**20**
Contratar	5	15

Las cuotas igualadas combinan estos dos atributos al satisfacer tanto el saldo de tasas de falsos negativos como el saldo de tasas de falsos positivos.

En el ejemplo de reclutamiento, la solución no cumple con los criterios, ya que los solicitantes masculinos no calificados tienen muchas más probabilidades de ser contratados que las solicitantes femeninas no calificadas Y las solicitantes femeninas calificadas tienen muchas más probabilidades de ser rechazadas que los solicitantes masculinos.

Wortman Vaughan y Wallace, 2022)

Proyecto | La IA en el reclutamiento

Consideraciones en torno a las métricas de equidad

- Un sistema no puede lograr la paridad predictiva, el equilibrio de tasas de falsos positivos y el equilibrio de tasas de falsos negativos a la vez debido a restricciones matemáticas.
- Las compensaciones entre los diferentes objetivos de equidad son inherentes y deben gestionarse con cuidado.
- Las consideraciones de equidad deben contextualizarse dentro de la aplicación específica a la que se aplican.
- No todos los aspectos de la equidad son cuantificables, reconociendo la complejidad sociotécnica de la equidad.

Proyecto | La IA en el reclutamiento

Equidad individual

U
n

Licenciatura
1 año de experiencia laboral relacionada

B

Master
1 año de experiencia laboral relacionada

C

Master
Sin experiencia laboral relacionada

- La equidad individual difiere de los criterios de equidad basados en el grupo (las métricas discutidas anteriormente) al centrarse en el trato a los individuos.
- Este concepto afirma que los individuos con características similares deben recibir resultados comparables.
- Determinar una métrica para medir la similitud individual es complejo.
- Considere tres candidatos para un puesto de trabajo; ¿Es A más capaz de hacer el trabajo que C?

Zhong, 2018

FUENTE DE LAS IMÁGENES Elaboracion propia

Proyecto | La IA en el reclutamiento

IMPLEMENTACIÓN DEL MODELO

Informar a los candidatos sobre el uso de la IA en la contratación

Tras cuidadosas evaluaciones internas, la empresa está lista para lanzar su herramienta de selección selectiva en los puestos vacantes para recopilar los comentarios iniciales de los usuarios y solicitantes.

Es crucial que los solicitantes sean conscientes de que están interactuando con un sistema de contratación impulsado por la IA, lo que fomenta la confianza. La concienciación sobre las ventajas de las herramientas de IA debe establecerse desde el principio, lo que mejora la disposición de los solicitantes a utilizar tecnologías interactivas como los chatbots. Además, perfeccionar los procesos de contratación de IA y garantizar la claridad sobre el funcionamiento del sistema son clave para su adopción y éxito entre los candidatos.

Además, debe asegurarse de que haya un humano que supervise e intervenga cuando sea necesario.

Chen, 2023

Proyecto | La IA en el reclutamiento

Informar a los candidatos sobre el uso de la IA en la contratación | Ejemplos de herramientas

Evaluación del impacto ético de la UNESCO

Un ejemplo de la herramienta de evaluación del impacto ético de la Unesco presentada anteriormente

10.2.1. System Awareness:

10.2.1.1. Are users made fully aware when they are interacting with an AI system, as opposed to a human being?

10.2.1.2. Are individuals (directly or indirectly) impacted by the AI system made fully aware of when a decision (that impacted them) was informed by or made on the basis of an AI system or AI algorithms?

10.2.1.2.1. Are they made aware of the extent to which they are impacted, including the rationale, benefits and limitations of the decision(s)?

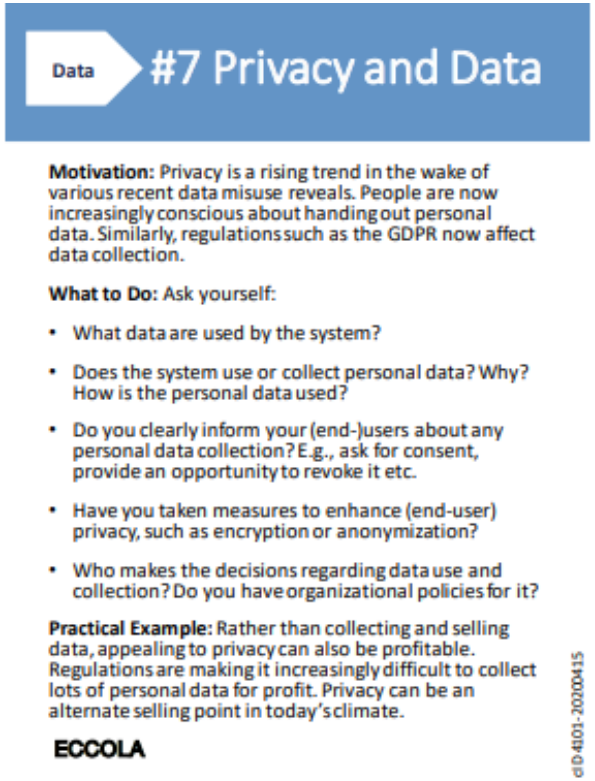
10.2.1.3. Have appropriate explanations been put in place to help users and other impacted individuals understand the decision-making process or how the system works when required?

10.2.1.4. Have appropriate explanations been put in place to help the government bodies in charge of regulation understand the decision-making process or how the system works when required?

10.2.1.5. Has the decision to adopt the AI system been documented and communicated online?

10.2.1.6. Can the AI system make any decisions which the physical persons or legal entities in charge of the system lack expertise or competence to critique, modify or override?

Tarjetas Eccola; un método para implementar sistemas de IA alineados éticamente



Data **#7 Privacy and Data**

Motivation: Privacy is a rising trend in the wake of various recent data misuse reveals. People are now increasingly conscious about handing out personal data. Similarly, regulations such as the GDPR now affect data collection.

What to Do: Ask yourself:

- What data are used by the system?
- Does the system use or collect personal data? Why? How is the personal data used?
- Do you clearly inform your (end-)users about any personal data collection? E.g., ask for consent, provide an opportunity to revoke it etc.
- Have you taken measures to enhance (end-user) privacy, such as encryption or anonymization?
- Who makes the decisions regarding data use and collection? Do you have organizational policies for it?

Practical Example: Rather than collecting and selling data, appealing to privacy can also be profitable. Regulations are making it increasingly difficult to collect lots of personal data for profit. Privacy can be an alternate selling point in today's climate.

ECCOLA

cID 4101-20200415

Unesco, 2023; Vakkuri et al. 2021

Proyecto | La IA en el reclutamiento

OPERACIÓN Y MONITOREO

Después de la implementación

Si bien se realizaron evaluaciones internas de equidad antes del lanzamiento, la empresa también debe estar preparada para auditorías externas independientes. También es ventajoso buscar continuamente comentarios después de la implementación y establecer canales de participación continuos para las partes interesadas. Esto implica mantener a los trabajadores informados e involucrados en consultas y procesos participativos durante toda la implementación del sistema de IA dentro de la organización.

IA confiable de HLEG

GRACIAS

Cofinanciado por la Unión Europea

Financiado por la Unión Europea. Las opiniones y puntos de vista expresados solo comprometen a su(s) autor(es) y no reflejan necesariamente los de la Unión Europea o los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA pueden ser considerados responsables de ellos.

2022-1-ES01-KA220-HED-000085257