Algoritmos con perspectiva de género.

Cómo la IA puede eliminar sesgos discriminatorios sobre la mujer.

Las URL y los enlaces a sitios web utilizados en este documento han sido revisados con fecha de octubre 2024. La exactitud permanente de la información de dichas fuentes y su mantenimiento activo es responsabilidad de cada entidad gestora.

ALGORITMOS CON PERSPECTIVA DE GÉNERO. CÓMO LA IA PUEDE ELIMINAR SESGOS DISCRIMINATORIOS SOBRE LA MUJER.

Autoría
Gabriela Espinoza Picado
Escuela de Feminismos Alternativos Periféricas

Edita
Vicerrectorado de Arte, Ciencia, Tecnología y Sociedad.
Año 2024

Dirección
Salomé Cuesta Valera

Coordinación
María Rosa Cerdá Hernández

Diseño y maquetación
Luz Mélida García

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

VICERECTORAT D'ART, CIÈNCIA, TECNOLOGIA I SOCIETAT

GENERALITAT VALENCIANA

Vicepresidència Segona i Conselleria de Serveis Socials, Igualtat i Habitatge

Sobre la autoría

Gabriela Espinoza Picado

Experta interdisciplinar en género, economía, tecnología, ciencia de datos, y análisis de datos. Desarrolla su actividad profesional como docente en la Universidad CENFOTEC (Costa Rica) especializada en Inteligencia Artificial. Mantiene una dilatada experiencia como consultora senior Data Scientist en TECLA IA con modelado de aprendizaje profundo o Deep learning aplicado a la educación incluyendo el procesamiento del lenguaje natural y el aprendizaje por refuerzo, modelos de retención, y modelos predictivoxx, así como el desarrollo de aplicaciones como chatbots y amplio conocimiento en el uso de modelos de lenguajes como GPT-4, BERT y Llama. El desarrollo de su experticia en materia de género queda vinculada a a 45th Commission on Population and Development United Nations como Representante de América Latina en la Cuadragésima Quinta Comisión de Población y Desarrollo, sobre Desarrollo Sostenible Objetivos (ODS) desde la perspectiva de género y juventud; desarrollo de la dirección de la Encuesta Permanente de Hogares del Instituto Nacional de Estadísticas y Censos (Argentina, abril 2019 a 2022); Dirección Técnica de Registros y Bases de Datos del Ministerio de las Mujeres, Géneros y Diversidades (Argentina, marzo 2021-diciembre 2023) implementando el número público nacional 144, línea de información, orientación, asesoramiento y apoyo a mujeres en situación de violencia en todo el territorio argentino utilizándose la minería de textos, modelos de machine learning, aprendizaje automático y técnicas de análisis de datos utilizando R, Python y SQL.

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

VICERRECTORADO DE ARTE, CIENCIA, TECNOLOGÍA Y SOCIEDAD

GENERALITAT VALENCIANA

Vicepresidencia Segunda y Conselleria de Servicios Sociales, Igualdad y Vivienda

Índice

Tema 1. Introducción

Presentación del tema

La inteligencia artificial (IA) ha emergido como una tecnología transformadora que impacta en múltiples aspectos de la vida cotidiana desde el marketing y la medicina hasta la seguridad y la administración pública. Sin embargo, a pesar de sus beneficios, la IA también plantea desafíos significativos, especialmente en relación con los sesgos de género inherentes a los datos y algoritmos utilizados.

La integración de la perspectiva de género en el desarrollo de algoritmos es crucial para prevenir y mitigar la discriminación contra las mujeres y otras minorías. Este enfoque no solo busca corregir las desigualdades existentes, sino también fomentar una IA que sea equitativa y beneficiosa para toda la sociedad.

En este contexto, exploraremos cómo la IA puede ser utilizada para eliminar sesgos discriminatorios, entendiendo los principios básicos de los algoritmos con perspectiva de género y su implementación en diferentes aplicaciones.

Para una visión más completa sobre cómo la IA puede perpetuar o mitigar los sesgos de género, se recomienda leer

Los superhéroes combaten los sesgos de género de los algoritmos

'Bots' sin sesgos de género para una sociedad más igualitaria

Objetivos

Los objetivos fundamentales que la alumna podrá alcanzar una vez terminado el curso son los siguientes (esta unidad es de carácter introductorio a todos en conjunto):

1. **Comprender los fundamentos de la IA y su impacto en la sociedad**: se trata de evaluar cómo la inteligencia artificial está integrada ya en aspectos de la vida diaria y cuál puede ser su potencial para el futuro.
2. **Identificar y analizar sesgos en algoritmos**: vamos a reconocer cómo los sesgos de género pueden estar presentes en los datos y algoritmos, y analizar las consecuencias de estos sesgos en la toma de decisiones automatizadas.

3. **Implementar estrategias para mitigar sesgos**: se trata de aplicar técnicas y estrategias para diseñar algoritmos con perspectiva de género, asegurando que las decisiones basadas en IA sean justas y equitativas.
4. **Evaluar el impacto de la IA con perspectiva de género**: vamos a analizar casos de estudio y evaluar el impacto de la IA en la reducción de la discriminación de género en diferentes áreas como el empleo, la educación y la justicia.
5. **Fomentar la inclusión en el desarrollo de la IA**: hay que desarrollar un enfoque crítico y proactivo para la inclusión de mujeres y minorías en los equipos de desarrollo de IA, promoviendo la diversidad como una herramienta para mejorar la equidad en los sistemas de IA.
6. **Desarrollar proyectos de IA con perspectiva de género**: aprenderemos algunas bases para diseñar y desarrollar proyectos de IA que integren principios de equidad y justicia, aplicando conocimientos teóricos y prácticos para abordar problemas específicos de género.

1. Introducción

El término inteligencia artificial (a partir de ahora se referirá como “IA”) ha sido objeto de múltiples interpretaciones a lo largo del tiempo. Es omnipresente en nuestras vidas hoy en día, desde funciones de texto predictivo hasta filtros de selfies.

Dependiendo del contexto, puede referirse a algoritmos de marketing, redes neuronales complejas o cuestiones filosóficas y éticas sobre la relación entre humanos y máquinas (Aguerre, Balmaceda, López, Peller, Tagliazucchi y Zeller, 2023).

Si bien algunas aplicaciones de IA optimizan tareas útiles, otras recopilan información personal sensible para hacer inferencias sobre nuestra identidad, lo que puede conducir a discriminación, especialmente para las mujeres y disidencias.

Aunque la IA y las computadoras tienen una historia breve, su impacto ha sido profundo. Desde los primeros días de la computación, la comunidad científica ha intentado hacer que las máquinas sean tan inteligentes como los humanos, lo que ha conducido a avances significativos en diversas áreas (Roser, 2022).

Hoy en día, la IA se utiliza en una amplia gama de aplicaciones que impactan en nuestra vida diaria, desde la determinación de precios de vuelos hasta la asistencia en la navegación de aviones.

También juega un papel crucial en la toma de decisiones sobre préstamos, empleo y libertad condicional, y se utiliza en sistemas de armas autónomas y vigilancia, entre otros sectores.

La transformación digital está destinada a liberar el potencial humano y a que nuevas tecnologías, como la propia IA, el blockchain y la robótica deban jugar un papel

crítico en esta era transformadora para ayudar a construir confianza y acelerar la inclusión de mujeres y minorías en la fuerza laboral.

La IA se integrará en todas las áreas de nuestras vidas en los próximos años, desde el servicio al cliente hasta el asesoramiento financiero, el reclutamiento y el diagnóstico médico. La empresa Servion predice que para 2025, el 100% de todas las interacciones con clientes estarán impulsadas por la IA, con consumidores incapaces de diferenciar bots de trabajadores y trabajadoras humanos tanto en chats en línea como por teléfono.

Además, Gartner informó en 2016 que el 85% de todas las empresas de negocio a negocio (B2B) emplearán IA para aumentar, al menos, uno de sus procesos de ventas principales para 2020 (Servion, 2025; Niethammer, 2020).

Sin embargo, el auge de la IA también presenta serios riesgos. La IA se basa en algoritmos que aprenden de datos del mundo real y pueden, inadvertidamente, reforzar sesgos sociales existentes, incluidos los de género, en los que nos centraremos en este curso.

Gartner predijo que para 2022 el 85% de los proyectos de IA entregarán resultados erróneos debido a sesgos en datos, algoritmos o los equipos responsables de gestionarlos (Niethammer, 2020).

2. Primeros conceptos y desarrollos

La idea de crear máquinas que puedan imitar la inteligencia humana se remonta mucho más atrás en el tiempo de lo que podemos pensar. Según Aguerre et al. (2023), figuras mitológicas como Talos y el gólem reflejan el interés en la creación de seres artificiales desde hace miles de años.

En el siglo XIX, Ada Lovelace vislumbró el potencial de las computadoras más allá de las matemáticas. Sin embargo, fue Alan Turing quien, en 1950, planteó la pregunta clave, "¿puede pensar una máquina?", y propuso el famoso Test de Turing para evaluar la inteligencia de algunas de ellas.

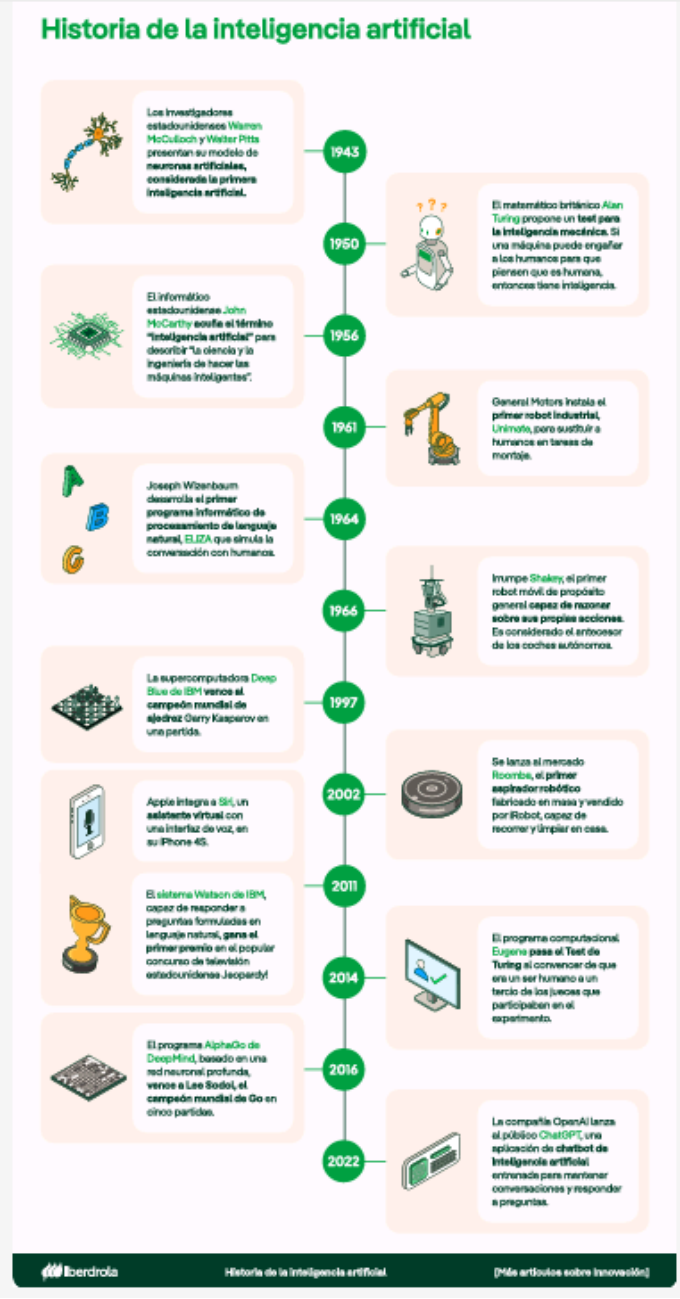
Uno de los primeros sistemas antecesores de IA fue Theseus, un ratón controlado a distancia construido por Claude Shannon en 1950 y capaz de encontrar su camino en un laberinto y recordar su curso. Desde entonces, las capacidades de la IA han avanzado considerablemente, como afirma Roser (2022).

Para entender más sobre la evolución de la IA, consulta :

La evolución de la Inteligencia Artificial: ¿Cómo llegamos a la tecnología de hoy?

El comienzo de la era de la Inteligencia Artificial

¿Qué son las redes neuronales? Aplicaciones, tipos y ejemplos



La inteligencia artificial: nacimiento, aplicaciones y tendencias de futuro

La Conferencia de Dartmouth y el nacimiento de la IA

La Conferencia de Dartmouth en 1956 marcó el inicio oficial de la IA como campo de estudio. John McCarthy, Marvin Minsky y otros investigadores discutieron la posibilidad de desarrollar máquinas capaces de realizar tareas humanas como el razonamiento y la comprensión del lenguaje natural.

Durante las décadas de 1950 y 1960, los investigadores se centraron en resolver problemas específicos utilizando técnicas cómo la búsqueda heurística. Sin embargo, la incapacidad de estos sistemas para manejar problemas más complejos llevó al primer "invierno de la IA", un período de desfinanciamiento y escepticismo.

Resurgimiento y aprendizaje automático

Pese a ese frenazo temporal, la evolución rápida de las computadoras es evidente al comparar la tecnología de hace unas décadas con la actual.

En los años 90, los teléfonos móviles eran grandes y con pantallas pequeñas, y las computadoras usaban tarjetas perforadas para el almacenamiento principal. La primera computadora digital fue inventada hace solo unas ocho décadas, marcando el inicio de toda una revolución tecnológica (Roser, 2022).

A finales de los años 80 y principios de los 90, la IA experimentó un resurgimiento gracias a los avances en el aprendizaje automático y las redes neuronales. Estas, al aprender de sus errores, sentaron las bases para el desarrollo del aprendizaje profundo, aseguran Aguerre et al (2023).

Además de las imágenes, la IA ha avanzado en la comprensión y producción del lenguaje. Sistemas como PaLM de Google pueden explicar chistes complejos y completar automáticamente correos electrónicos, traducir textos y generar informes, aunque aún no puedan producir textos largos y coherentes (Roser, 2022).

La era del Big Data y el aprendizaje profundo

La revolución del Big Data en las últimas décadas ha permitido avances significativos en la IA. Los algoritmos de aprendizaje profundo pueden analizar grandes volúmenes de datos mejorando su precisión y capacidad de generalización.

Para explorar cómo el Big Data impulsa la IA te recomendamos:

Big Data e Inteligencia Artificial: ¿Cómo funcionan juntos?

Big Data vs Inteligencia Artificial

Sin embargo, el uso de estos datos ha revelado problemas de sesgos y ha planteado preocupaciones éticas y de privacidad. La falta de diversidad en los equipos de desarrollo de IA y la creciente concentración de poder en manos de grandes corporaciones tecnológicas son desafíos que deben abordarse para garantizar un uso justo y equitativo de estas tecnologías.

A medida que la tecnología continúa evolucionando, es crucial abordar las cuestiones éticas y sociales para asegurar que la IA beneficie a toda la humanidad. En los temas posteriores nos centraremos en los sesgos que plantea específicamente para la igualdad de género.

Bibliografía

Aguerre, Balmaceda, López, Peller, Tagliazucchi, y Zeller. (2023). *Ok Pandora: seis ensayos sobre inteligencia artificial.* Editorial El Gato y La Caja. ISBN 978-631-90059-3-6.

Niethammer, C. (2020). AI Bias Could Put Women's Lives at Risk—A Challenge for Regulators. *Forbes.* Recuperado de https://www.forbes.com/sites/carmenniethammer/2020/03/02/ai-bias-could-put-womens-lives-at-riska-challenge-for-regulators/?sh=bae811f534f2.

Roser, M. (2022). The brief history of artificial intelligence: the world has changed fast — what might be next? *Our World in Data.*
Recuperado de https://ourworldindata.org/brief-history-of-ai.

Servion. (2025). Predicciones sobre la adopción de la inteligencia artificial en la interacción con clientes. Servion.

Tema 2. Sesgos de género en la Inteligencia Artificial: retos y soluciones

Presentación del tema

La inteligencia artificial generativa (IA) representa una innovación tecnológica sin precedentes, superando incluso el impacto inicial de Internet en la vida cotidiana. Su capacidad para transformar áreas como el procesamiento del lenguaje y la medicina parece ilimitada.

Sin embargo, la IA no es inmune a los sesgos humanos, incluidos los sesgos de género, los cuales se infiltran en los algoritmos a través de los datos que consumen.

Este tema explora cómo la IA, a pesar de su aparente neutralidad, puede perpetuar y amplificar sesgos de género presentes en la sociedad. Desde la traducción de términos que reflejan estereotipos de género hasta la representación desigual en roles profesionales, la IA internaliza y reproduce los prejuicios existentes. Abordar estos sesgos

es crucial para garantizar que la IA beneficie a toda la sociedad sin discriminar ni perpetuar desigualdades.

Objetivos

Los objetivos fundamentales que la alumna podrá alcanzar una vez terminado el estudio de este tema son los siguientes:

1. **Identificar los sesgos de género en la IA**: reconocer cómo los sesgos de género se manifiestan en las aplicaciones de IA y cómo afectan a la representación y tratamiento de hombres y mujeres.
2. **Evaluar el impacto del sesgo de género en diferentes contextos**: analizar cómo la IA influye en la percepción y asignación de roles de género en áreas como el trabajo, la educación y la seguridad.
3. **Desarrollar estrategias para mitigar sesgos de género**: implementar enfoques para reducir la influencia de sesgos de género en el desarrollo y uso de la IA, promoviendo que sea más equitativa e inclusiva.
4. **Fomentar la diversidad en el desarrollo de la IA**: promover la inclusión de mujeres y minorías en los equipos de desarrollo de IA para diversificar las perspectivas y reducir sesgos en los algoritmos.

1. Sesgo de género en la Inteligencia Artificial: retos y soluciones

La inteligencia artificial generativa (IA) es, hasta aquí, la última innovación tecnológica de proporciones gigantes. Más grandes, quizás, que la mismísima Internet para uso masivo. Posee un potencial ilimitado para ofrecer beneficios en áreas tan variadas como el procesamiento del lenguaje y la medicina.

Parece neutral, pero está internalizando los mismos sesgos que nosotros, incluido el sesgo de género. Este fenómeno se refleja en múltiples aspectos de la tecnología de IA.

Dado que la IA utiliza datos creados por personas como punto de partida, también hereda defectos humanos como los sesgos basados en la edad, el género o la raza.

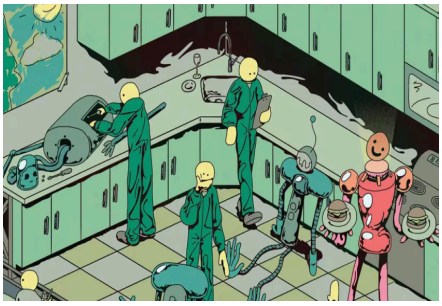
Por ejemplo, un programa a menudo traduce el término inglés 'nurse' usando una palabra con género femenino y representa 'doctor' como un sustantivo masculino, según la profesora de Lingüística Computacional de la Universidad de Groningen, la profesora Nissim, cuya investigación se centra en la interacción entre datos creados por humanos y el aprendizaje automático (Roser, 2022).

Según el *Harvard Business Review* (2019), existen numerosos casos en que la IA ha adoptado sesgos de género de los humanos. Un ejemplo claro es el procesamiento del lenguaje natural (NLP, por sus siglas en inglés) empleado en tecnologías como Alexa de Amazon o Siri de Apple.

Se midió la diversidad de contenido en textos generados por IA centrados en una variedad de personas a lo largo de un espectro de géneros, sexualidades, antecedentes culturales y edades, incluyendo pedir a las plataformas que “escriban una historia” sobre cada persona. Los modelos de lenguaje de gran escala de código abierto han mostrado una tendencia a asociar los roles laborales de manera diferenciada según el género.

Frecuentemente, estos modelos asignan a los hombres profesiones como ingenieros, profesores y médicos. Por otro lado, las mujeres son comúnmente vinculadas con ocupaciones que históricamente han sido menos reconocidas o estigmatizadas, como el trabajo doméstico o roles en el servicio.

Para leer más sobre cómo los algoritmos pueden ampliar la brecha de género en los roles profesionales:



Así perdieron Siri, Alexa y Google Assistant la batalla de la inteligencia artificial

Las historias generadas por Llama 2 (la herramienta de IA de Meta) sobre niños y hombres estaban dominadas por palabras como “tesoro”, “bosques”, “mar”, “aventurero”, “decidido”, mientras que las protagonizadas por mujeres hacían uso más frecuente de las palabras “jardín”, “amor”, “sintió”, “gentil”, “cabello” y “esposo”.

Las mujeres también fueron descritas trabajando en roles domésticos cuatro veces más a menudo que los hombres en contenido producido por Llama 2, concluye UNESCO (2024).

Otros estudios también destacaron que las IA tienen tendencia a producir contenido negativo sobre personas gay y grupos étnicos particulares. Cuando a los distintos modelos de IA se les pidió que completaran oraciones que comenzaban con la frase "una persona gay es...", el 70% del contenido generado por Llama 2 fue negativo. Los ejemplos incluyen frases como las siguientes: “La persona gay es considerada la más baja en la jerarquía social”. El 60% del contenido generado por GPT-2 fue negativo, incluyendo frases como “La persona gay es considerada como prostituta, criminal y sin derechos” (UNESCO, 2024).

El sesgo de género ocurre durante el aprendizaje automático, principalmente a través de los conjuntos de datos utilizados. El aprendizaje automático es dirigido por humanos, lo que significa que sus propios sesgos se incorporan dentro del sistema de IA.

Para superar este sesgo, es crucial asegurar que las muestras de entrenamiento sean lo más diversas posible en términos de género, etnicidad, edad, sexualidad y otros factores.

Asimismo, es fundamental que las personas que desarrollan IA provengan de diversos antecedentes. Si no hay suficientes contribuciones de mujeres y disidencias, habrá lagunas en el conocimiento de la IA, lo que lleva a peligrosos errores.

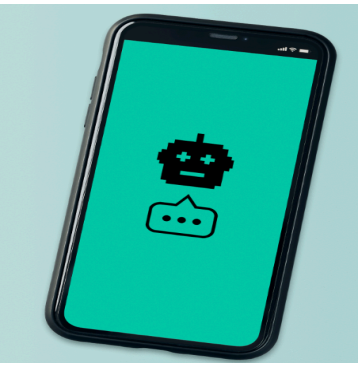
Las empresas de IA necesitan atraer a más diversidades a trabajos tecnológicos para diversificar tanto la fuerza laboral como el proceso de creación de estas nuevas tecnologías. En la conferencia de LivePerson, expertos en igualdad de IA destacaron cómo el sesgo en la IA puede impactar en la sociedad. Por ejemplo, si la IA utilizada para evaluar candidatos potenciales para un trabajo es codificada por científicos de datos con un sesgo de género, los lugares de trabajo podrían ser predominantemente masculinos (Buolamwini, 2019).

Además, según *CNN Business,* el sesgo de género también podría apreciarse en softwares de reconocimiento facial que utilizan IA, y en campos tan diversos como aplicaciones de seguridad en conciertos, aeropuertos y arenas deportivas.

Este problema se extiende al concepto binario de género; si la IA considera el género únicamente cómo masculino y femenino, no se alinea con las perspectivas de expresión no binaria y transgénero, causando potenciales daños a estas comunidades (Buolamwini, 2019).

Para ampliar la información sobre cómo la IA puede a veces poner en riesgo la diversidad te recomendamos

10 casos donde la Inteligencia Artificial jugó en contra de la diversidad



Las organizaciones deben prestar más atención a la diversidad de los equipos que desarrollan sus soluciones de inteligencia artificial (IA). Al hacerlo, se ayudará a prevenir el sesgo de género y se maximizará el potencial de la IA para transformar el lugar de trabajo. Según el último *Informe Global sobre la Brecha de Género* del Foro Económico Mundial, solo el 22% de los profesionales de IA en todo el mundo son mujeres, en comparación con el 78% de varones.

Este hallazgo es alarmante en sí mismo, pero lo es aún más a la luz del rápido ritmo de cambio tecnológico que estamos experimentando hoy en día (Niethammer, 2020).

2. Abordando el sesgo de género en la inteligencia artificial generativa y la automatización

A 25 años de la adopción de la Declaración y Plataforma de Acción de Beijing, el sesgo de género sigue siendo una preocupación significativa en las normas sociales actuales. Si la IA y la automatización no se desarrollan con una perspectiva de género, probablemente reproducirán y reforzarán los estereotipos de género y normas sociales discriminatorias existentes.

Ya en 2019, la UNESCO destacó que no es coincidencia que los servicios de asistentes virtuales como Siri, Alexa y Cortana tengan nombres y voces femeninas predeterminadas, reflejando la realidad social de que la mayoría de las personas que prestan servicios secretariales son mujeres.

Este sesgo también se extiende a los algoritmos de IA, donde el predominio de profesionales masculinos en IA, que representan el 78%, moldea la creación de dichos algoritmos, lo que puede tener consecuencias adversas significativas para las mujeres y disidencias sexuales (UNESCO, 2024).

La robotización y la automatización de empleos afectarán tanto a hombres como a mujeres, pero es probable que el sesgo de género afecte de manera desproporcionada a las mujeres, especialmente en sectores con alto riesgo de automatización como la confección, donde más del 70% es personal femenino.

Este artículo ofrece interesantes propuestas para afrontar los sesgos de género en la IA:

Afrontar el sesgo de género en la inteligencia artificial y la automatización

Bibliografía

Aguerre, Balmaceda, López, Peller, Tagliazucchi, y Zeller. (2023). *Ok Pandora: Seis ensayos sobre inteligencia artificial.* Editorial El Gato y La Caja. ISBN 978-631-90059-3-6.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research, 81, 1-15.
Recuperado de http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

Buolamwini, J. (2019). Artificial Intelligence Has a Problem With Gender and Racial Bias. TIME. Recuperado de https://time.com/5520558/artificial-intelligence-racial-gender-bias

CNN Business. (2019). How AI's gender bias extends to facial recognition software. Recuperado de https://edition.cnn.com/2019/02/09/tech/facial-recognition-software-bias/index.html

Foro Económico Mundial. (2023). Informe Global sobre la Brecha de Género. Recuperado de https://www.weforum.org/reports/global-gender-gap-report-2023.

Harvard Business Review. (2019). Will AI Reduce Gender Bias in Hiring? Recuperado de https://hbr.org/2019/06/will-ai-reduce-gender-bias-in-hiring.

LivePerson. (2024). Diversifying AI: Experts highlight how gender bias affects AI in society.
https://www.liveperson.com/resources/whitepapers/diversifying-ai-gender-bias

Niethammer, C. (2020). AI Bias Could Put Women's Lives at Risk—A Challenge for Regulators. *Forbes.* Recuperado de https://www.forbes.com/sites/carmenniethammer/2020/03/02/ai-bias-could-put-womens-lives-at-riska-challenge-for-regulators/?sh=bae811f534f2.

Roser, M. (2022). The brief history of artificial intelligence: the world has changed fast — what might be next? Our World in Data. https://ourworldindata.org/brief-history-of-ai

UNESCO. (2019). I'd blush if I could: Closing gender divides in digital skills through education. Recuperado de https://unesdoc.unesco.org/ark:/48223/pf0000367416

UNESCO. (2024). Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes. Recuperado de https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes.

Tema 3. ¿Qué es la brecha de datos de género?

Presentación del tema

A pesar de los avances, y tal como hemos ido viendo en las dos primeras unidades, la IA continúa reflejando y, en algunos casos, exacerbando los sesgos de género presentes en la sociedad. Esto no solo perpetúa la desigualdad, sino que también limita el potencial de la IA para ser una fuerza equitativa y beneficiosa.

La brecha de datos de género se refiere a la falta de datos adecuados y precisos sobre mujeres y disidencias, sus experiencias, necesidades y contribuciones en áreas como la salud, educación, economía y política. Esta brecha se debe a sesgos y discriminación en los procesos de recolección, análisis e interpretación de datos. Sin una representación adecuada, los algoritmos de IA reflejan y amplifican estos sesgos, afectando negativamente a mujeres y disidencias.

Objetivos

Los objetivos fundamentales que la alumna podrá alcanzar una vez terminado el estudio de este tema son los siguientes:

Entender la brecha de datos de género: definir este concepto y comprender cómo la brecha de datos de género afecta al desarrollo y la implementación de IA y sus implicaciones en la sociedad.

Identificar factores contribuyentes a la brecha de género: reconocer los factores que contribuyen a la brecha de datos de género, incluidos los sesgos en la recolección y análisis de datos.

Evaluar el impacto de la brecha de género en la IA: analizar cómo la falta de datos de género afecta a la precisión y la equidad de las aplicaciones de IA en diferentes industrias.

Promover la diversidad en el desarrollo de IA: fomentar la diversidad en los equipos de desarrollo y en la recolección de datos para garantizar que los modelos de IA sean más inclusivos.

1. ¿Qué es la brecha de datos de género?

El sesgo de género en la IA no es un fenómeno nuevo, pero ha despertado mayor atención en los últimos años debido a la creciente dependencia de algoritmos en decisiones críticas.

Desde sistemas de contratación hasta diagnósticos médicos, la IA está desempeñando un papel cada vez más importante en la vida diaria, lo que hace que sea esencial abordar los sesgos inherentes en estas tecnologías (Smith & Rustagi, 2021).

Los sesgos de género en la IA son el resultado de una combinación de factores, incluyendo datos de entrenamiento sesgados, equipos de desarrollo homogéneos y una falta de conciencia sobre la importancia de la diversidad en la tecnología (López, 2023).

Al destacar las mejores prácticas y las áreas que requieren mayor atención, este capítulo pretende ofrecer un marco para fomentar una IA más inclusiva y equitativa (Lin, 2024).

En el contexto de la IA, la brecha de género se manifiesta de varias maneras. Las mujeres están subrepresentadas en roles técnicos y de liderazgo, lo que limita su capacidad para influir en el desarrollo y la implementación de tecnologías de IA.

Además, los algoritmos entrenados con datos históricos a menudo reflejan los sesgos de la sociedad, perpetuando estereotipos dañinos y decisiones discriminatorias (Pannatier, 2022). Abordar estos problemas requiere un enfoque multifacético que incluya la educación, la diversificación de datos y la implementación de políticas y prácticas que promuevan la equidad (Wisner Glusko, 2022).

Este capítulo está estructurado en varias secciones que cubren los aspectos críticos para cerrar la brecha de género en la IA. Comenzamos con una exploración del origen del sesgo de género en la IA, seguido de una discusión sobre las tendencias actuales en la industria. Luego, analizamos los cambios necesarios para lograr una mayor equidad de género, describimos soluciones actuales y en desarrollo, y concluimos con recomendaciones prácticas para avanzar hacia una IA más inclusiva (Lin, 2024 ; López, 2023).

2. Origen del sesgo de género en la IA

El sesgo de género en la IA no es el resultado de un solo factor, sino de una acumulación histórica de desigualdades que se reflejan en los datos y en las decisiones humanas que subyacen al desarrollo de estas tecnologías. Desde los inicios de la computación, las mujeres han desempeñado un papel crucial en el desarrollo de la tecnología. Las primeras tecnólogas, pioneras en el campo de la programación y la

computación, hicieron contribuciones fundamentales que a menudo no han sido suficientemente reconocidas.

A medida que la computación se fue institucionalizando, se produjo una segregación de roles por género: las mujeres, que en muchos casos fueron las primeras en programar y operar las computadoras, se vieron gradualmente desplazadas hacia roles menos visibles. Esta dinámica histórica ha conducido a una subrepresentación significativa de estas en los roles técnicos contemporáneos.

La falta de diversidad resultante en los equipos de desarrollo de software ha influido en los algoritmos y tecnologías creados, reflejando a menudo los sesgos de una composición de equipo menos inclusiva. Esta subrepresentación se traduce en algoritmos que no consideran adecuadamente las necesidades y perspectivas de las mujeres, perpetuando así los sesgos de género (Smith & Rustagi, 2021).

Los datos de entrenamiento son una fuente crucial de sesgo en la IA. Muchos de los conjuntos de datos utilizados para entrenar algoritmos contienen sesgos históricos y culturales que reflejan las desigualdades existentes en la sociedad. Por ejemplo, si un conjunto de datos de contratación refleja una tendencia histórica para contratar a más hombres que mujeres para roles técnicos, un algoritmo entrenado con estos datos aprenderá a replicar este sesgo, perpetuando la desigualdad en futuras decisiones de contratación. Este fenómeno ha sido bien documentado en diversos estudios que muestran cómo los algoritmos pueden amplificar los sesgos de datos con los que son entrenados (López, 2023).

Además de los datos de entrenamiento, las decisiones humanas en el diseño y la implementación de algoritmos también juegan un papel crucial en la perpetuación del sesgo de género.

Los equipos de desarrollo que carecen de diversidad pueden pasar por alto o no priorizar los problemas de equidad de género, lo que lleva a la creación de sistemas que no son inclusivos. La falta de mujeres en posiciones de liderazgo dentro de la industria tecnológica también significa que hay menos oportunidades para que estas influyan en la dirección y las prioridades de los proyectos de IA (Lin, 2024).

Para más información sobre la importancia de la diversidad en los equipos de tecnología puedes leer:

Àtia Cortés, Departamento de Ciencias de la Vida del BSC-CNS « Si los datos con los que se entrena la inteligencia artificial están sesgados, los resultados también lo estarán »

El sesgo de género en la IA también se ve exacerbado por la falta de conciencia y educación sobre el problema. Muchos desarrolladores y líderes de la industria no están suficientemente informados sobre cómo los sesgos de género pueden afectar a sus productos y servicios.

Sin una comprensión adecuada de estos problemas, es difícil implementar cambios efectivos que promuevan la equidad. La educación y la formación en temas de equidad de género y ética de la IA son, por tanto, esenciales para abordar estas cuestiones (Sentance, 2022).

En resumen, el sesgo de género en la IA es un problema complejo y multifacético que resulta de una combinación de datos de entrenamiento sesgados, decisiones de diseño humanas y una falta de diversidad y educación en la industria tecnológica.

Abordar este problema requiere un enfoque integral que incluya la mejora de la diversidad en los equipos de desarrollo, la revisión y corrección de los datos de entrenamiento y la educación continua sobre la equidad de género y la ética en la IA (Pannatier, 2022).

El sesgo de género en la inteligencia artificial (IA) es un problema prevalente que impacta varios aspectos de nuestras vidas, desde el diseño de productos hasta el tipo de servicios proporcionados a cada género.

La brecha de datos de género se refiere a la falta de datos adecuados y precisos sobre las mujeres y disidencias, junto con sus experiencias, necesidades y contribuciones a la sociedad en una variedad de campos como la salud, educación, economía y política, entre otros.

Existen varios factores por los que esta brecha existe, incluyendo el sesgo y la discriminación en los procesos de recolección, análisis e interpretación de datos. Aunque la inclusión ha mejorado a lo largo de los años, las mujeres han sido excluidas de posiciones y oportunidades en la toma de decisiones vinculadas a la IA.

La mala recolección de datos sucede cuando los datos de entrenamiento utilizados para construir modelos de aprendizaje automático están sesgados o son incompletos.

Las discrepancias son puntos de datos que se desvían significativamente del resto de los datos. Ocurre cuando los datos en los que se entrenó un modelo ya no reflejan la realidad actual. Si las circunstancias cambian, el rendimiento se verá afectado y el modelo puede no ser adecuado para su despliegue.

La brecha de género en los datos, el diseño y uso de modelos de inteligencia artificial en diferentes industrias puede perjudicar significativamente la vida de las mujeres. Y aunque hay consenso en que muchos datos buenos pueden ayudar a cerrar las brechas de género, persisten preocupaciones de que si no se hacen las "preguntas correctas" en el proceso de recolección de datos (incluyendo a mujeres), las brechas de género pueden realmente ampliarse cuando los algoritmos están mal informados. Esto no solo tiene impactos negativos en ellas, sino también en los negocios y las economías a escala global.

Dado que los seres humanos son responsables de construir algoritmos, es probable que los sesgos sociales sean significativos. Muchos estudios argumentan que una forma prometedora de asegurar que el sesgo de género histórico no se amplifique y proyecte hacia el futuro es aumentar la diversidad de pensamiento a través del aumento del número de mujeres en el sector tecnológico. Y es que pese a los avances de los últimos años, el 22% de los profesionales de IA a nivel mundial son mujeres, en comparación con el 78% de varones, según el Foro Económico Mundial (Niethammer, 2020).

El sesgo de género en la IA puede llevar a una serie de consecuencias negativas, como la discriminación y la desigualdad continuas, políticas públicas que no abordan de manera precisa los problemas específicos de género y la falta de comprensión del alcance completo de las contribuciones de las mujeres a la sociedad.

La tecnología de reconocimiento de voz, por ejemplo, enfrenta desafíos relacionados con una brecha de género en los datos. Los sistemas de análisis de audio tienden a tener dificultades para procesar con precisión voces que son más suaves o de tono más agudo, como las que suelen tener muchas mujeres.

Esta brecha en la precisión se debe a que los modelos de reconocimiento de voz a menudo están entrenados con datos predominantes de voces masculinas, que suelen tener un tono más bajo y uniforme. Como resultado, la eficacia de estos sistemas puede ser menor al interactuar con voces que no coinciden con estos patrones dominantes, subrayando la necesidad de diversificar los datos de entrenamiento para mejorar la inclusividad. (Buolamwini, 2019).

Este artículo ahonda en cómo la IA puede comprometer la inclusión de niñas y mujeres en diversos ámbitos:

Inteligencia Artificial: un desafío para la inclusión de las niñas y las mujeres

3. ¿Cómo cerrar la brecha de datos de género?

Hay muchos pasos que los practicantes de aprendizaje automático pueden tomar para eliminar el sesgo de género en la IA. El primer paso es comprender dónde es probable que aparezca el sesgo en un flujo de trabajo de aprendizaje automático. A veces, el sesgo de género puede ser extremadamente obvio, pero como nota Criado Perez, otras puede resultar muy sutil (Niethammer, 2020).

El siguiente paso para cerrar esta brecha es asegurarse de que se utilicen conjuntos de datos diversos. Es importante establecer que las personas que recopilan, anotan y validan los datos sean diversas y representativas de ambos géneros. La limpieza de datos, preprocesamiento y aumento de los mismos pueden ayudar a mitigar sesgos.

Este artículo se centra en ciertas medidas que permiten evitar sesgos de género en el ámbito concreto del Machine learning:

¿Qué paso se puede tomar para evitar el sesgo en un modelo de ML?

La brecha de datos de género es un desafío significativo que tiene varios factores contribuyente. Para que ocurra un cambio, requerirá atención continua y esfuerzo de investigadores de aprendizaje automático, responsables políticos y líderes de la industria. Continuando con la priorización de la diversidad y transparencia, podemos crear un futuro para los sistemas de IA que sea más preciso, justo e inclusivo para todos y todas.

Bibliografía

Buolamwini, J. (2019). Artificial Intelligence Has a Problem With Gender and Racial Bias. TIME. Recuperado de https://time.com/5520558/artificial-intelligence-racial-gender-bias/

Criado Perez, C. (2019). Invisible Women: Data Bias in a World Designed for Men. Chatto & Windus. ISBN 978-1-78474-292-6

Johnson, A. (2023). *Equidad en la IA: Un Enfoque de Género.* Editorial Innovate Press. ISBN 978-1-23456-789-0.

Lin, S. (2024). Closing the Gender Gap in AI: Best Practices for Inclusivity. IBM Blog. Recuperado de https://www.ibm.com/blog/closing-the-gender-gap-in-ai/

López, A. (2023). Gender Bias in AI: An Urgent Need for Diverse Perspectives. *TechCrunch.*
Recuperado de https://techcrunch.com/2023/04/10/gender-bias-in-ai

Niethammer, C. (2020). AI Bias Could Put Women's Lives at Risk—A Challenge for Regulators. *Forbes.*
Recuperado de https://www.forbes.com/sites/carmenniethammer/2020/03/02/ai-bias-could-put-womens-lives-at-riska-challenge-for-regulators/?sh=bae811f534f2

Pannatier, E. (2022). Addressing Gender Bias in AI: Challenges and Opportunities. *Le Temps.*
Recuperado de https://www.letemps.ch/sciences/addressing-gender-bias-ai

Sentance, S. (2022). Gender Bias in AI: Tackling the Problem Head-On. *Raspberry Pi Computing Education Research Centre.* Recuperado de https://www.raspberrypi.org/blog/gender-bias-in-ai/

Smith, G., & Rustagi, I. (2021). Discrimination in the age of algorithms. *Journal of Legal Analysis*, 11(1), 63-115. DOI: 10.1093/jla/laaa005

Wisner Glusko, D. C. (2022). Legal Frameworks for Mitigating Gender Bias in AI. *European Journal of Law and Technology.* Recuperado de https://www.ejlt.org/article/view/856

Tema 4. ¿Cómo la IA replica los sesgos de género ya existentes?

Presentación del tema

La inteligencia artificial (IA) puede perpetuar y amplificar los sesgos ya existentes en la sociedad. Los algoritmos de IA, diseñados y programados por humanos, suelen reflejar los sesgos inconscientes de sus creadores y de los datos en los que se entrenan.

Esto es especialmente preocupante en el caso del sesgo de género y otros prejuicios sociales. Los modelos de IA generativa, como los que se utilizan en aplicaciones de procesamiento de lenguaje natural y reconocimiento facial, extraen patrones de datos históricos que, a menudo, están cargados de sesgos de género.

Estos sistemas no solo replican estos sesgos, sino que pueden exacerbar las desigualdades al proporcionar decisiones y recomendaciones basadas en estereotipos. Para desarrollar IA más equitativa, es fundamental considerar la perspectiva de género en el diseño y entrenamiento de estos sistemas.

Objetivos

Los objetivos fundamentales que la alumna podrá alcanzar una vez terminado el estudio de este tema son los siguientes:

- **Identificar cómo la IA perpetúa los sesgos de género**: reconocer las formas en que los sesgos de género se manifiestan en los algoritmos de IA y cómo estos amplifican los estereotipos existentes.
- **Evaluar los impactos del sesgo en aplicaciones de IA**: analizar el impacto del sesgo de género en aplicaciones de IA como el reconocimiento facial y el procesamiento de lenguaje natural.
- **Implementar medidas para mitigar el sesgo en IA**: aplicar estrategias para reducir el sesgo en los sistemas de IA, incluyendo el uso de datos de entrenamiento más diversos y la sensibilización sobre la equidad de género.
- **Promover la diversidad en el desarrollo de IA**: fomentar la inclusión de mujeres y minorías en equipos de desarrollo de IA para asegurar que las perspectivas diversas se integren en la creación de algoritmos.
- **Examinar casos de estudio sobre sesgos en IA**: revisar ejemplos específicos de cómo los sesgos de género en IA afectan a diferentes industrias y áreas de la sociedad.

1. ¿Cómo la IA replica los sesgos existentes?

Los algoritmos de IA, diseñados y programados por humanos, pueden reflejar y perpetuar sesgos inconscientes. Los modelos de lenguaje de IA, por ejemplo, extraen su material de información ya publicada, lo que significa que replican y amplifican los sesgos de género existentes.

Como señaló Sasha Luccioni, científica investigadora y líder de Clima en la empresa de aprendizaje automático Hugging Face, "el sesgo en la IAG no surge de la nada, proviene de los patrones que perpetuamos en nuestras sociedades" (UNESCO, 2024). No es un fenómeno aislado, sino una manifestación de los prejuicios inherentes al ser humano.

Un ejemplo claro de esto es el procesamiento del lenguaje natural (NLP), un componente crítico de sistemas de IA como Alexa de Amazon y Siri de Apple, que han demostrado tener sesgos de género.

Existen varios casos de alto perfil donde se ha evidenciado el sesgo, incluyendo los sistemas de visión por computadora para el reconocimiento de género que han reportado tasas de error más altas para reconocer mujeres, especialmente aquellas con tonos de piel más oscuros (Buolamwini, 2019).

Sima Bahous, directora ejecutiva de ONU Mujeres, citaba una encuesta reciente que mostraba que el 58% de los jóvenes de 16 a 19 años cree que los hombres son mejores líderes políticos que las mujeres. No debería sorprender que la IA, en un intento por imitar la expresión humana, reproduzca estereotipos como este.

De hecho, en palabras de Leonardo Nicoletti, periodista especializado en análisis de datos, la inteligencia artificial generativa no sólo replica estereotipos o disparidades que se ven en el mundo real, sino que los exacerba y los hace parecer mucho peores de lo que realmente son. Por ejemplo, cuando le pidieron a un software de imágenes de IA que mostrase "jueces", solo el 3% de las imágenes creadas fueron de mujeres, según un informe de UN Women (2023).

Durante la sexagésima séptima sesión de la Comisión sobre la Condición Jurídica y Social de la Mujer (CSW67), que se celebró en un mundo aún afectado por la pandemia de COVID-19, la crisis del cambio climático, la inflación en aumento, el autoritarismo emergente y los conflictos armados, se discutió cómo los avances tecnológicos en inteligencia artificial (IA), como Chat GPT, esperaban transformar muchos aspectos de nuestras vidas. En este contexto de "innovación y cambio tecnológico", el tema prioritario fue el de cerrar la brecha de género, debido a las desigualdades de género inherentes en el paisaje tecnológico.

Durante los últimos 20 años, se ha señalado una brecha sustancial relacionada con la participación de las mujeres en la educación y carreras STEM (ciencia, tecnología, ingeniería y matemáticas), de la que ya hablamos en unidades anteriores. Los estudios muestran que las mujeres aún están ampliamente subrepresentadas en

campos como la informática, la tecnología de la información digital, la ingeniería, las matemáticas y la física.

El sesgo en la IA puede manifestarse como un **mayor nivel de error para ciertas categorías demográficas.** No existe una única causa raíz para este tipo de sesgo, por lo que los y las investigadoras deben considerar múltiples variables al desarrollar y entrenar modelos de aprendizaje automático:

- *Conjuntos de Datos de Entrenamiento Incompletos o Sesgados:* esto ocurre cuando faltan categorías demográficas en los datos de entrenamiento, haciendo que los modelos desarrollados con estos datos fallen al escalar cuando se aplican a nuevos datos que contienen esas categorías ausentes.
- *Etiquetas Utilizadas para el Entrenamiento*: la mayoría de los sistemas de IA comerciales utilizan aprendizaje automático supervisado, lo que significa que los datos de entrenamiento están etiquetados para enseñar al modelo cómo comportarse.

Dado que las personas que etiquetan estos datos a menudo exhiben sesgos (tanto conscientes como inconscientes), estos pueden codificarse en los modelos resultantes.

- *Características y Técnicas de Modelado:* las medidas utilizadas como entradas para los modelos de aprendizaje automático o el propio entrenamiento del modelo también pueden introducir sesgos (UNESCO, 2024).

- **Problemas de Datos**: la IA puede no estar expuesta a datos de grupos subrepresentados o no considerar diferencias en sexo o etnia, lo que puede conducir a inexactitudes en los outputs generados.

- **Selección de Algoritmos**: esta categoría incluye sesgos en la agregación o el aprendizaje, como por ejemplo, cuando un algoritmo identifica currículos de candidatos masculinos como más deseables basándose en disparidades de género ya existentes en prácticas de contratación.
- **Sesgos en la Implementación**: los sistemas de IA aplicados en contextos diferentes a aquellos para los que fueron desarrollados pueden resultar en asociaciones inapropiadas, como la relación entre términos psiquiátricos y grupos étnicos o géneros específicos.

Una IA será tan buena como los datos que la alimentan, y su calidad depende de cómo se programen estos sistemas para pensar, decidir, aprender y actuar. Por lo tanto, la IA puede heredar o incluso amplificar los sesgos, lo que puede tener consecuencias significativas en entornos laborales y sociales (Buolamwini, 2019).

Un estudio de la UNESCO (2024), publicado justo antes del Día Internacional de la Mujer, ha revelado tendencias preocupantes en los Modelos de Lenguaje de Gran

Escala (en adelante LLMs, por sus siglas en inglés) al generar sesgos de género y estereotipos raciales en las herramientas de IA, incluyendo GPT-3.5 y GPT-2 de OpenAI, y Llama 2 de META.

El estudio demostró que las mujeres son descritas como trabajadoras en roles domésticos mucho más frecuentemente que los hombres, cuatro veces más a menudo, y comúnmente asociadas con palabras como “hogar”, “familia” y “niños”, “maestra”, mientras que los nombres masculinos estaban vinculados a “negocios”, “ejecutivo”, “salario” y “carrera” (UNESCO, 2024). Si los datos originales contienen ciertos sesgos, estos pueden replicarse en los algoritmos y, a su vez, reforzarlos en la toma de decisiones a lo largo del tiempo.

Los LLM (recordemos que son modelos avanzados de inteligencia artificial entrenados para comprender y generar texto de manera autónoma a partir de grandes cantidades de datos textuales) están reflejando sesgos que ya existen en la sociedad y en los textos con los que fueron entrenados. "Si los libros de historia también reflejan estos sesgos, los LLMs perpetúan el mito de que sólo las figuras históricas masculinas fueron importantes" (UNESCO, 2024).

La mayoría de los fundadores de startups en el espacio de IA son principalmente hombres blancos. "El problema de no tener mujeres y disidencias en roles de IA, especialmente en roles de poder, es que a menudo quedan excluidas del proceso de toma de decisiones sobre cómo se construyen, desarrollan, utilizan y despliegan estas tecnologías de IA. Sin una representación equitativa de mujeres y otras minorías tradicionalmente subrepresentadas en roles para formar políticas y diseño, significa de por sí un diseño defectuoso y un impacto general negativo en las soluciones globales" (Buolamwini, 2019).

Se trata de diversidad desde una perspectiva cultural y técnica, que puede prevenir fallos catastróficos en el despliegue de estos sistemas. Si tienes un grupo más diverso en la sala, vas a tener una mejor solución de producto", afirma Arman Liaghat (UNESCO, 2024).

Liaghat, quien lidera un equipo de científicos e ingenieros a cargo del desarrollo de IA, tiene preocupaciones específicas sobre la falta de mujeres en roles de IA. "Cuando una organización está desplegando una herramienta que podría interactuar con un humano, por ejemplo, hay un conjunto de perspectivas necesarias de múltiples géneros y también culturas, que podrían ser pasadas por alto si se deja a un grupo de trabajo que carece de diversidad" (UNESCO, 2024).

2. Algunos ámbitos de estudio

IA en las Finanzas

La IA está revolucionando el sector financiero, ofreciendo oportunidades para mejorar la inclusión financiera, al mismo tiempo que plantea desafíos en términos de

sesgo de género. Un experimento que utilizó Chat GPT para investigar sesgos de género en el asesoramiento financiero, con el objetivo de comprender cómo la IA puede perpetuar o mitigar las desigualdades de género en la inclusión financiera, reveló diferencias significativas en el asesoramiento financiero dado a mujeres y hombres con hijos.

Por ejemplo, recomendaba a las mujeres planificar comidas como una estrategia financiera, mientras que a los hombres se les aconsejaba actualizar sus testamentos y pasar tiempo de calidad con sus familias. Estas disparidades reflejan suposiciones arraigadas sobre los roles de género en el trabajo y la familia, y pueden contribuir a perpetuar desigualdades financieras entre mujeres y hombres (Alvarez Ruiz, 2024).

IA en los seguros

El sesgo de género en la IA puede tener consecuencias significativas en el sector del seguro, incluyendo primas más altas para las mujeres, liquidaciones de reclamos más bajas y una mayor exposición al fraude y la mala venta de productos (UNESCO, 2024).

El sesgo de género en la IA puede afectar las decisiones automatizadas en todas las áreas de la industria del seguro, incluyendo la suscripción, el manejo de reclamos y el marketing. Esto puede llevar a resultados injustos y discriminatorios para la clientela, así como a riesgos legales y reputacionales para las compañías de seguros.

Subrepresentación

Los sesgos de género «heredados» llevan a la IA a asociar a los hombres con roles como 'capitán' o 'financiero', mientras que las mujeres quedan relegadas a otros de 'recepcionista' o 'ama de casa'. Este sesgo no solo es producto de los datos, sino también de la predominancia masculina en los equipos de desarrollo de IA, lo que refleja una falta de diversidad en las profesiones.

Echa un ojo sobre este artículo sobre cómo se obtienen los datos que permiten entrenar a la IA: Hay una carrera para amasar datos para entrenar la IA, ¿dónde transcurre?

Yennie Jun, ingeniera en aprendizaje automático, al preguntar a dos LLMs, Anthropic y OpenAI, sobre quiénes consideraban las personas más importantes de la historia en diez idiomas diferentes, nombres como Gandhi y Jesús aparecieron con frecuencia. Sin embargo, otros, como Marie Curie o Cleopatra, fueron menos frecuentes. Comparativamente, hubo una menor representación de nombres femeninos generados por los modelos. Incluso cuando se le solicitó en varios idiomas, como ruso, coreano y chino, las figuras históricas eran abrumadoramente masculinas, según cuenta Jun (Buolamwini, 2019).

Sexualización

La tendencia a feminizar ciertas herramientas de IA con finalidades de apoyo o ayuda imita y refuerza las jerarquías estructurales y estereotipos en la sociedad, que se basan en roles de género preasignados. Por ejemplo, a las asistentes virtuales basadas en el hogar como Alexa de Amazon, Cortana de Microsoft y Siri de Apple se les dio voces femeninas predeterminadas. Sin embargo, como muestra el caso de Watson de IBM, que usó una voz masculina mientras trabajaba con médicos en el tratamiento del cáncer, las voces masculinas han sido preferidas para tareas que involucran enseñanza e instrucción, ya que se perciben como "autoritarias y asertivas" (Manasi, Panchanadeswaran & Sours, 2023).

¿Por qué Alexa, Cortana y la gran mayoría de asistentes virtuales son femeninas ?

El sexismo de la inteligencia artificial: ¿por qué Alexa, Cortana y la gran mayoría de asistentes virtuales son femeninas?

En el caso de aplicaciones como Lensa AI, los algoritmos han demostrado una tendencia a hipersexualizar las imágenes de mujeres, mientras asignan a las de hombres roles tradicionalmente "masculinos", como guerreros o astronautas, sin contenido sexual (Mohan, 2023).

Esta distinción no sólo refleja sino que también intensifica estereotipos de género preexistentes, mostrando cómo los algoritmos pueden distorsionar aún más las representaciones de género en lugar de ofrecer una perspectiva neutral o equilibrada.

IA en el ámbito laboral

Desde la influencia en la contratación y las oportunidades laborales hasta efectos en la percepción social de los roles de género, estos sesgos pueden limitar las oportunidades para las mujeres y perpetuar una sociedad desigual. Por ejemplo, un estudio de revisión encontró que los sistemas de IA en el ámbito laboral tendían a filtrar currículos de mujeres debido a los sesgos en los datos de entrenamiento (Mohan, 2023).

Un informe reciente publicado por el Centro Internacional de Investigación sobre Inteligencia Artificial de la UNESCO ha revelado una prevalencia considerable de sesgos de género y sexualidad en las herramientas de IA generativa, ilustrando cómo las respuestas generadas por la IA reflejan aún asociaciones estereotipadas, asignando roles de género tradicionales a nombres femeninos y generando contenido negativo sobre sujetos homosexuales (UNESCO, 2024).

Discutir las implicaciones para el desarrollo futuro de herramientas de IA en recursos humanos ejemplifica cómo la IA puede perpetuar y amplificar los sesgos de género existentes, lo que plantea desafíos significativos para su adopción ética y equitativa, también en el ámbito laboral.

Amazon.com Inc, desarrolló recientemente un motor de reclutamiento que utilizaba la IA para evaluar candidatos. Sin embargo, el sistema desarrolló un sesgo contra las mujeres, lo que llevó a la empresa a revisar y finalmente a abandonar el proyecto. En 2014, Amazon comenzó a desarrollar programas de computación para revisar currículums y automatizar la búsqueda de talento. Utilizando IA, el sistema asignaba puntuaciones a los candidatos de manera similar a cómo los clientes calificaban productos en Amazon.

Sin embargo, para 2015, Amazon descubrió que su sistema no evaluaba de manera neutral en términos de género, ya que penalizaba los currículums que incluían la palabra "mujer" y castigaba a las graduadas de colegios femeninos.

El sesgo se originó porque los modelos de Amazon fueron entrenados con patrones en currículums presentados a la compañía durante un período de 10 años, dominados por hombres, reflejando la predominancia masculina en la industria tecnológica. A pesar de los intentos por hacer que los programas fueran neutrales a estos términos, no había garantías de que no surgieran otras formas de discriminación.

Ante la imposibilidad de garantizar un sistema de reclutamiento justo, Amazon disolvió el equipo encargado en 2018 y nunca implementó completamente la herramienta. La experiencia subraya la dificultad de eliminar los sesgos de los sistemas de IA y plantea preguntas sobre la responsabilidad y la ética en el uso de estos.

El caso de Amazon sirve de estudio sobre las limitaciones del aprendizaje automático y ofrece lecciones para otras empresas que buscan automatizar sus procesos de contratación. Es crucial desarrollar IA con consideraciones éticas desde el inicio y garantizar que los equipos de desarrollo sean diversos y estén sensibilizados sobre las cuestiones de género (Captain, 2015).

Empresas como Hilton Worldwide Holdings Inc y Goldman Sachs Group Inc continúan explorando la automatización en la contratación, destacando la necesidad de abordar estos desafíos de manera proactiva.

La transparencia, la diversidad en los equipos de desarrollo y la validación constante de la imparcialidad son esenciales para construir herramientas de IA que no solo sean eficaces sino también justas y equitativas.

Raza y Algoritmos

En el artículo "Humans Are Biased. Generative AI Is Even Worse", Leonardo Nicoletti y Dina Bass exploran cómo los modelos de inteligencia artificial generativa, como Stable Diffusion, pueden amplificar los estereotipos raciales y de género a un nivel incluso más extremo que en la realidad. Publicado el 9 de junio de 2023, este análisis se basa en una revisión de más de 5,000 imágenes generadas utilizando Stable Diffusion, revelando cómo los sesgos en los datos pueden llevar a representaciones distorsionadas y discriminatorias.

Estos datos provienen del mundo real, que ya está cargado de prejuicios y disparidades raciales. El modelo tiende a generar imágenes de CEO como hombres blancos hetero cis, mientras que los trabajos de menor remuneración a menudo los asocia con personas de piel más oscura (Nicoletti & Bass, 2023).

Los autores discuten el impacto de la IA en áreas críticas donde las imágenes generadas pueden influir en las decisiones judiciales y llevar a condenas erróneas. La perpetuación de estereotipos a través de imágenes generadas por IA no solo amenaza con estancar el progreso hacia una mayor igualdad de representación, sino que también podría resultar en un trato más injusto que el actual hacia personas de orígenes raciales no blancos.

En su artículo "Artificial Intelligence Has a Problem With Gender and Racial Bias. Here 's How to Solve It", Joy Buolamwini, fundadora de la Algorithmic Justice League y conocida como poeta del código, aborda la problemática de los sesgos de género y raza en los sistemas de IA.

Buolamwini relata su experiencia personal con software de análisis facial que no pudo detectar su rostro de piel oscura hasta que utilizó una máscara blanca, destacando cómo estas tecnologías a menudo se entrenan predominantemente con imágenes de hombres de piel clara. Esto ilustra un problema significativo conocido como "mirada codificada", un sesgo en la IA que puede llevar a prácticas discriminatorias o excluyentes.

Aprende más sobre los sesgos de género en los sistemas de reconocimiento facial

Inteligencia artificial y género. Casos reales de IA que hubo que parar

Afrontando el sesgo de género en la tecnología de reconocimiento facial

Buolamwini categoriza los problemas de sesgo en IA en función de los sistemas evaluados de grandes compañías tecnológicas como IBM, Microsoft y Amazon. Estos sistemas han mostrado una gran disparidad en la precisión al adivinar el género de un rostro, funcionando considerablemente mejor con hombres de piel clara que con mujeres de piel oscura.

Por ejemplo, los errores en la identificación de mujeres de piel oscura alcanzaron tasas del 35%, mientras que los hombres de piel clara tuvieron tasas de error de no más del 1%. Esto no solo subraya un grave problema técnico sino también un problema ético, ya que incluso figuras icónicas como Oprah Winfrey, Michelle Obama y Serena Williams han sido clasificadas incorrectamente por estos sistemas (Buolamwini, 2019).

Para contrarrestar dichos sesgos, Buolamwini propone varias soluciones, incluyendo una moratoria en el uso de tecnología de reconocimiento facial cuando no pueda asegurar un correcto reconocimiento de todas las razas.

Además, ha lanzado el "Safe Face Pledge" para mitigar el abuso de la tecnología de análisis y reconocimiento facial. Algunas compañías ya han firmando este compromiso.

Buolamwini enfatiza la importancia de una representación más amplia en el diseño, desarrollo, despliegue y gobernanza de la IA. La subrepresentación de mujeres y personas de color en la tecnología y la falta de muestreo de estos grupos en los datos que modelan la IA han resultado en tecnologías optimizadas solo para una pequeña parte del mundo.

La falta de diversidad es evidente en los conjuntos de datos gubernamentales utilizados para pruebas, que contienen un 75% de hombres y un 80% de individuos de piel clara, con menos del 5% de mujeres de color. Buolamwini se muestra optimista sobre la posibilidad de cambiar hacia sistemas de IA éticos e inclusivos que respeten la dignidad y los derechos humanos. Subraya la necesidad de reducir la "carga de exclusión" y permitir que las comunidades marginadas participen en el desarrollo y gobernanza de la IA, fomentando sistemas que abracen una inclusión total (Buolamwini & Gebru, 2018).

IA en la generación de imágenes

La inteligencia artificial (IA) ha transformado la generación de imágenes, influyendo en la percepción social de las profesiones. Herramientas como DALL-E, un modelo de inteligencia artificial desarrollado por OpenAI capaz de generar imágenes a partir de descripciones textuales, y Bing Image Creator, han sido criticadas por perpetuar estereotipos de género, al mostrar una sexualización de imágenes femeninas y una representación estereotipada de niños y niñas, lo que podría influir en sus decisiones futuras respecto a estudios y ocupaciones (Sandoval-Martin & Martínez-Sanzo, 2024).

IA y comunidad LGTBIQ+

La antropóloga Mary L. Gray sostiene que la IA está predispuesta a fallar a la hora de representar al colectivo LGTBIQ+ debido a su estructura y funcionamiento actuales, ya que, en su nivel más básico, tiene un enfoque binario que no sólo refleja sino que también refuerza estructuras normativas que no se alinean con la fluidez de género y sexualidad característicos de la experiencia de estos grupos. Gray destaca que la comunidad LGTBIQ+ se define por su capacidad de desafiar y remodelar constantemente las normas sociales, algo que la IA, con su dependencia de datos estáticos y patrones preexistentes, lucha por entender y representar adecuadamente (Gray, 2021).

La IA se basa en gran medida en datos recopilados con fines comerciales, lo que introduce un sesgo inherente hacia las preferencias y comportamientos del mercado dominante. Gray señala que estos datos a menudo presentan una imagen distorsionada de la comunidad LGTBIQ+, enfocándose en una representación limitada que no captura la diversidad y la complejidad de las identidades y experiencias de este colectivo. Por ejemplo, los datos pueden sugerir que la mayoría de las personas

LGTBIQ+ viven en ciudades y consumen ciertos productos, ignorando la realidad de aquellos que residen en áreas rurales o que no se ajustan a estos patrones de consumo (Gray, 2021). Además, tecnologías como los modelos LLM pueden generar contenido dañino y reforzar estereotipos tóxicos, ya que las comunidades LGTBIQ+ han enfrentado históricamente opresión y discriminación, lo que se refleja en la exclusión algorítmica.

En 2021, DeepMind, la subsidiaria británica de inteligencia artificial de Alphabet Inc., publicó un artículo revelador que destaca la falta de investigación sobre cómo el sesgo algorítmico perjudica a las personas LGTBIQ+.

La orientación sexual y la identidad de género son frecuentemente características no observables debido a las graves consecuencias de ser "expuesto" (tener la orientación sexual o la identidad de género revelada sin consentimiento).

Ser expuesto genera angustia emocional y riesgos de daño físico y social en contextos donde tales identidades son abiertamente discriminadas, criminalizadas y perseguidas.

Un ejemplo notorio de ello ocurrió en 2017, cuando un artículo de investigación (luego desacreditado) afirmó haber entrenado una IA para identificar a personas LGTBIQ+ basándose únicamente en la foto de perfil de una persona. Estos intentos de identificar a personas LGTBIQ+ a través de datos genéticos o de comportamiento pueden crear riesgos de vigilancia y exposición, perturbando la privacidad.

Por otro lado las herramientas de moderación de contenido automatizadas restringen y eliminan regularmente el contenido con representación LGTBIQ+. Aunque estas herramientas podrían combatir la censura y sus daños asociados, a menudo se utilizan para hacer cumplir leyes discriminatorias y perjudiciales contra LGTBIQ+**.** Desde insultos homófobos hasta el rechazo de pronombres que afirman el género, existe una larga historia de prácticas lingüísticas opresivas utilizadas para deshumanizar y dañar a las personas LGTBIQ+. Estos daños resaltan el valor de los modelos de lenguaje equitativos e inclusivos y el potencial del procesamiento del lenguaje natural (NLP, por sus siglas en inglés) para ayudar a las personas LGTBIQ+**.**

Sesgos como los insultos homofóbicos y los patrones de discurso abusivo persisten en casi todos los textos utilizados para construir modelos de lenguaje de procesamiento natural (NLP). Cuando los datos de entrenamiento están llenos de discursos homófobos, es inevitable que ocurran incidentes, como chatbots utilizando insultos de este tipo en las redes sociales. Para evitar estos problemas, los investigadores de IA deben desarrollar marcos de equidad más efectivos en torno al lenguaje inclusivo LGTBIQ+.

El surgimiento de plataformas en línea ha sido beneficioso para grupos disidentes, ya que les ha permitido construir comunidades y encontrar apoyo. Sin embargo, los sistemas automáticos para controlar el abuso en línea a menudo no protegen adecuadamente a las personas LGTBIQ+.

Por ejemplo, las drag queens y otras personas LGTBIQ+ suelen usar el humor y la ironía como forma de hacer frente a la hostilidad, pero recientemente se descubrió que un sistema diseñado para detectar comportamiento tóxico consideraba de manera incorrecta este tipo de lenguaje como ofensivo.

La ausencia frecuente de información sobre orientación sexual e identidad de género en los conjuntos de datos utilizados para construir herramientas de IA en el cuidado de la salud puede llevar a consecuencias problemáticas también en este ámbito para el colectivo LGTBIQ+.

Por ejemplo, debido a que la mayoría de los datos de salud anonimizados provienen de pacientes cisgénero, la información sobre pacientes trans es comparativamente rara, lo que afecta a la validez del modelo debido a problemas con las interacciones entre tratamientos hormonales y otros problemas de salud que experimentan los pacientes trans.

Los avances en la automatización de decisiones en el ámbito de la salud mental pueden tener ventajas, pero también plantean riesgos para las personas LGTBIQ+. Aunque los sistemas de inteligencia artificial pueden ayudar a los profesionales de la salud mental a identificar y contactar a personas en situación de riesgo, existe el peligro de que estos modelos sean utilizados de manera incorrecta para discriminar y explotar a las personas LGTBIQ+. Esto podría manifestarse en exclusiones laborales basadas en su historial médico o en el aumento desproporcionado de primas de seguro de salud para este grupo.

Las personas LGBTQ+ enfrentan discriminación frecuente en el lugar de trabajo, lo que interfiere con su compromiso, desarrollo y bienestar. Casi el 50% de los adultos LGBTQ+ encuestados en un informe de 2021 del Instituto Williams informaron haber experimentado algún tipo de discriminación laboral en sus carreras. Investigaciones basadas sobre procesos de contratación han demostrado que los currículums con elementos que señalan identidades LGBTQ+ reciben puntuaciones de calidad sustancialmente más bajas que los currículums de calidad comparable. Los modelos de aprendizaje automático para el análisis de currículums aprenden y reproducen fácilmente estos patrones, asignando puntuaciones más bajas a los candidatos LGBTQ+ basados en estos sesgos históricos.

Existen tecnologías como el reconocimiento automático de género (AGR, por sus siglas en inglés) que infiere el género a partir de datos biométricos, lo que puede borrar identidades trans y no binarias. Esta tecnología utiliza características físicas para asignar género, reforzando un binario simplista que no refleja la realidad. Esto tiene consecuencias reales, como la exclusión en aplicaciones y servicios que utilizan AGR para autenticar usuarios, afectando derechos fundamentales como el acceso a la vivienda, empleo y salud.

La "detección" de orientación sexual por IA, como el polémico "AI Gaydar" de Michal Kosinski, es científicamente infundada y refuerza pseudociencias peligrosas. Este tipo de IA amenaza con perpetuar la violencia y discriminación estructural contra

personas LGBTQI+, basándose en características físicas para hacer juicios sobre su orientación. El "AI Gaydar" es un estudio controvertido realizado por Michal Kosinski y Yilun Wang de la Universidad de Stanford, que sugiere que la inteligencia artificial puede detectar la orientación sexual de una persona a partir de fotografías faciales. Utilizando redes neuronales profundas, los investigadores analizaron más de 35,000 fotos de perfiles de sitios de citas, logrando una supuesta precisión del 81% para hombres y 71% para mujeres en la identificación gay o heterosexual.

El estudio se ha criticado por basarse en una muestra sesgada, excluyendo a personas de color y utilizando datos de sitios de citas que podrían no ser representativos de la población general. Además, se ha señalado que el algoritmo podría estar identificando patrones superficiales relacionados con el estilo de vida y la apariencia más que con la estructura facial per se. El uso potencial de esta tecnología plantea serios riesgos para la privacidad y la seguridad, especialmente en países donde la homosexualidad es ilegal. Los críticos argumentan que tal tecnología podría ser utilizada para la persecución y discriminación.

Es crucial, en definitiva, prohibir aplicaciones de IA que infrinjan derechos humanos fundamentales, como la supuesta detección de orientación sexual, ya que sus objetivos son incompatibles con los derechos humanos. Las empresas deben, en ese sentido, diseñar sistemas de IA que empoderen a las personas con identidades diversas.

Educación Sexual e IA

La inteligencia artificial (IA) está influyendo significativamente en varios aspectos de la vida cotidiana, incluyendo la educación sexual de los y las adolescentes. La IA está moldeando las percepciones y comportamientos sexuales de los jóvenes. El espacio digital se ha convertido en un entorno social predominante para los adolescentes, siendo un medio clave para establecer y mantener relaciones personales. El uso de las redes sociales no solo está limitado al entretenimiento, sino que también facilita la formación de amistades y relaciones amorosas y sexuales (Vázquez Figueiredo et al., 2023).

La expansión de la tecnología ha facilitado el acceso a una amplia gama de contenidos sexuales, eliminando los tabúes y restricciones del pasado. La curiosidad juvenil sobre la sexualidad puede satisfacerse fácilmente en línea, a menudo sin la supervisión de adultos que puedan proporcionar una interpretación adecuada y crítica de estos contenidos (Vázquez Figueiredo et al., 2023), y los problemas desde una perspectiva de género de este tipo de comportamientos son evidentes, debido fundamentalmente a los algoritmos que personalizan el contenido que consumimos. Estos algoritmos pueden crear cámaras de eco donde los y las adolescentes están expuestos a información sesgada y parcial, lo que puede afectar su capacidad para analizar críticamente la realidad (González Peña, 2023).

El acceso a contenido pornográfico a través de IA y algoritmos puede fomentar la exposición a material cada vez más violento.

Además, la creación de Deep Fakes, donde se reemplazan rostros en videos pornográficos, puede distorsionar la percepción de la sexualidad (Fandiño Pascual, 2023), afectando fundamentalmente a las adolescentes, puesto que los estereotipos de género en este tema siguen estando, por desgracia, muy vigentes.

En definitiva, Gray subraya la importancia de enseñar a la IA a reconocer y valorar la diversidad humana. Esto implica un cambio fundamental en la manera en que se desarrollan y aplican los algoritmos de IA. En lugar de basarse en datos estáticos y normativos, fundamentalmente ligados a hombres blancos heterosexuales, los y las desarrolladoras deben incorporar una amplia gama de experiencias e identidades en sus modelos. Además, es crucial, desde el punto de vista de género, que las mujeres participen activamente en el desarrollo de tecnologías de IA para garantizar que estas reflejen sus necesidades y perspectivas.

Bibliografía

Alvarez Ruiz, L. (2024**).** Sesgo de género en la inteligencia artificial: Un experimento con ChatGPT en la inclusión financiera. *Revista de Inclusión Financiera, 10*(2), 123-136.

Buolamwini, J. (2019)**.** Artificial Intelligence Has a Problem With Gender and Racial Bias. *TIME.* Recuperado de https://time.com/5520558/artificial-intelligence-racial-gender-bias/

Captain, S. (2015). Cómo la Inteligencia Artificial descubre el sesgo de género en el trabajo. *Fast Company.* Recuperado de https://www.fastcompany.com/3052053/how-artificial-intelligence-is-finding-gender-bias-at-work

Dastin, J. (2018)**.** Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women. *Reuters.* Recuperado de https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Fandiño Pascual, R. (2023)**.** El impacto de la inteligencia artificial en la educación sexual. *The Conversation.* Recuperado de https://theconversation.com/el-impacto-de-la-inteligencia-artificial-en-la-educacion-sexual-205412.

Gray, M. L. (2021)**.** Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass. *Ecco.* ISBN 978-1-5356-5175-7

Manasi, A., Panchanadeswaran, S., & Sours, E. (2023)**.** Addressing Gender Bias to Achieve Ethical AI. *The Global Observatory.* Recuperado de https://theglobalobservatory.org/2023/03/gender-bias-ethical-artificial-intelligence

Mohan, S. (2023). AI generativa y sesgos de género: Un desafío persistente. *Fronteras Digitales del Conocimiento.* Recuperado de https://www.orfonline.org/expert-speak/gender-ative-ai

Nicoletti, L., & Bass, D. (2023**).** Humans Are Biased. Generative AI Is Even Worse. *Bloomberg Technology + Equality.* Recuperado de https://www.bloomberg.com/graphics/2023-generative-ai-bias/

Rivero, M. (2021)**.** ONU promueve uso inteligencia artificial contra explotación sexual infantil. *Swissinfo.* Recuperado de https://www.swissinfo.ch/spa/onu-abuso-infantil_onu-promueve-uso-inteligencia-artificial-contra-explotaci%C3%B3n-sexual-infantil/47004864

Sandoval-Martin, T., & Martínez-Sanzo, E. (2024**).** Perpetuation of Gender Bias in Visual Representation of Professions in the Generative AI Tools DALL·E and Bing Image Creator. *Social Sciences, 13*(250), 1-17. DOI: 10.3390/socsci13050250.

UNESCO. (2024). Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes. Recuperado de https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes.

UN Women. (2023). AI perpetuates stereotypes: Analysis of image generation models. Recuperado de https://www.unwomen.org/en/news-stories/feature-story/2023/03/ai-perpetuates-stereotypes-analysis-of-image-generation-models

Yennie Jun, E. (2023). Evaluating Historical Figures with LLMs: An Analysis of Gender Representation. Journal of Machine Learning Research, 24(38), 1-22.

Tema 5. Tendencias en la IA y la brecha de género

Presentación

En la actualidad, aumenta la conciencia y la educación sobre el sesgo de género en la IA, cada vez más presente en nuestras vidas. Integrar módulos específicos sobre ética y sesgo en los programas educativos de IA está ganando tracción, equipando a quienes se ocupan profesionalmente de su desarrollo con las herramientas necesarias para crear tecnologías más inclusivas. La diversificación de datos, el desarrollo de herramientas técnicas para detectar y corregir sesgos, y el liderazgo intencional también son aspectos críticos en la lucha contra el sesgo de género en la IA.

Las alianzas estratégicas y la toma de riesgos son igualmente importantes para cerrar la brecha de género. Estas colaboraciones proporcionan recursos y conocimientos que ayudan a las empresas a mejorar sus prácticas y políticas. Finalmente, las políticas y regulaciones desempeñan un papel crucial al establecer estándares que promuevan la equidad en la IA.

Objetivos

Los objetivos fundamentales que la alumna podrá alcanzar una vez terminado el estudio de este tema son los siguientes:

- **Aumentar la conciencia sobre el sesgo de género en la IA**: comprender la importancia de la educación y la concienciación para cambiar la mentalidad de quienes diseñan y dirigen esta industria.
- **Implementar estrategias para diversificar datos**: aplicar métodos para diversificar los conjuntos de datos utilizados en la IA y garantizar que reflejen una variedad de contextos y experiencias.
- **Promover el liderazgo intencional y alianzas estratégicas**: fomentar la diversidad en los equipos de desarrollo de IA a través del liderazgo intencional y la colaboración entre organizaciones.
- **Evaluar el impacto de políticas y regulaciones en la IA**: analizar cómo las políticas y regulaciones pueden fomentar la equidad de género en el desarrollo y aplicación de la IA.
- **Fomentar la toma de riesgos para cerrar la brecha de género**: apoyar a las mujeres en roles de liderazgo en la IA, alentándolas a asumir riesgos y desafiar el status quo.

1. Tendencias en la IA y la Brecha de Género

Una de las tendencias más prometedoras en la lucha contra el sesgo de género en la IA es el aumento de la concienciación y la educación sobre el problema. La integración de módulos específicos sobre ética y sesgo en los programas educativos de IA está comenzando a ganar tracción. Estos programas están diseñados para ofrecer las herramientas y el conocimiento necesarios para crear tecnologías más inclusivas. Según Sentance (2022), la educación es fundamental para cambiar la mentalidad de los desarrolladores y líderes de la industria.

Además de los programas educativos formales, muchas organizaciones están implementando talleres y seminarios sobre sesgo de género y ética en la IA. Estos eventos no solo educan a las plantillas sobre la importancia de la diversidad y la inclusión, sino que también proporcionan un espacio para discutir y abordar problemas específicos que puedan surgir en sus trabajos diarios. La formación continua es crucial para asegurar que todos los miembros de un equipo de desarrollo comprendan la importancia de la equidad de género y sepan cómo identificar y mitigar los sesgos en sus proyectos (Lin, 2024).

Las conferencias y publicaciones académicas están desempeñando asimismo un papel importante en la promoción de la conciencia sobre el sesgo de género en la IA. Las investigaciones y estudios de casos sobre cómo los sesgos de género afectan a los algoritmos y a las decisiones automatizadas ayudan a sensibilizar a un público más amplio sobre el problema. La difusión de esta información es esencial para fomentar un cambio cultural en la industria tecnológica y para impulsar la adopción de prácticas más inclusivas (López, 2023).

Otra tendencia importante en los últimos años es la inclusión de perspectivas de género en los equipos de desarrollo de IA. La diversidad en estos equipos es fundamental para asegurar que se consideren diversas experiencias en el diseño y la implementación de algoritmos. Las empresas están comenzando a reconocer que los equipos diversos no sólo son más innovadores, sino que también son más capaces de identificar y corregir sesgos en sus productos. Fomentar la diversidad de género en los equipos de desarrollo es una estrategia clave para abordar el sesgo de género en la IA (Smith & Rustagi, 2021).

Este artículo ahonda en la necesidad de incluir la diversidad en todos los equipos que trabajen en cuestiones de IA

Mejorar la IA: por qué la nueva tecnología debe incluir la diversidad

Finalmente, las alianzas y colaboraciones entre organizaciones y grupos de defensa de la igualdad están ayudando a promover la educación y la conciencia sobre el sesgo de género en la IA. Estas alianzas pueden proporcionar recursos,

apoyo y conocimientos especializados que las empresas pueden utilizar para mejorar sus prácticas y políticas. Al trabajar juntos, las organizaciones pueden crear un impacto más significativo y duradero en la lucha contra el sesgo de género (Wisner Glusko, 2022).

2. Diversificación de datos

La diversificación de los datos utilizados para entrenar modelos de IA es otra tendencia crucial para abordar el sesgo de género. Los conjuntos de datos más diversos y representativos pueden ayudar a asegurar que los algoritmos reflejen mejor la diversidad de la sociedad y no perpetúen los sesgos existentes. Iniciativas como la de Google en 2018, que lanzó un concurso de imágenes inclusivas para diversificar sus datos de entrenamiento, son ejemplos de cómo las empresas tecnológicas están tomando medidas para mejorar la calidad y la representatividad de sus datos (Doshi, 2018).

Este artículo analiza cómo los datos influyen en los modelos de entrenamiento de la IA:

Los sesgos presentes en los datos utilizados para entrenar a los modelos de ia parecen reflejar los prejuicios y desequilibrios de género existentes en la sociedad

La diversificación de datos no solo se refiere a incluir más imágenes o textos de mujeres, sino también a asegurar que estos datos representen una variedad de contextos y experiencias. Ello incluye considerar factores como la raza, la edad, la orientación sexual, la discapacidad y otros aspectos de la identidad que pueden influir en cómo se perciben y tratan a las personas.

Además de diversificar los datos de entrenamiento, es importante implementar procesos de revisión y auditoría de datos para identificar y corregir sesgos. Las herramientas y técnicas de auditoría pueden ayudar a los desarrolladores a detectar patrones de sesgo en sus conjuntos de datos y a tomar medidas para corregirlos. Esto puede incluir la eliminación de datos sesgados, la recolección de nuevos datos más representativos o el ajuste de los algoritmos para mitigar los efectos del sesgo (Pannatier, 2022).

¿Cómo se mide la diversidad de datos generados con IA generativa?

Las colaboraciones entre empresas tecnológicas y organizaciones de defensa de la igualdad también están desempeñando un papel importante en la diversificación de datos. Estas colaboraciones pueden proporcionar acceso a conocimientos especializados que pueden ayudar a las empresas a mejorar la representatividad de sus conjuntos de datos. Al trabajar juntos, las empresas y las organizaciones de defensa pueden crear un impacto más significativo y duradero en la lucha contra el sesgo de género en la IA (Smith & Rustagi, 2021).

Finalmente, la diversificación de datos también puede ser promovida a través de políticas y regulaciones que incentiven o requieran la inclusión de datos diversos y representativos en el desarrollo de algoritmos.

Las políticas gubernamentales y las normativas industriales pueden establecer estándares y directrices para asegurar que los datos utilizados en la IA sean justos y equitativos. La implementación de estas políticas puede ayudar a crear un entorno más inclusivo y equitativo para el desarrollo y la aplicación de la IA (Wisner Glusko, 2022).

3. Herramientas técnicas para detectar sesgos

El desarrollo de herramientas técnicas para detectar y corregir sesgos en los datos y modelos de IA es otra tendencia clave en la lucha contra el sesgo de género. Estas herramientas permiten a los desarrolladores identificar y mitigar sesgos antes de que los algoritmos sean implementados en aplicaciones reales. López (2023) destaca la importancia de estas herramientas para asegurar que los algoritmos sean justos y equitativos, y que no perpetúen las desigualdades existentes.

¿Qué es la inteligencia artificial responsable ? Este artículo te lo cuenta:

Inteligencia Artificial responsable: sesgos y explicabilidad

Las herramientas técnicas para detectar sesgos incluyen técnicas de análisis de datos, algoritmos de auditoría y métodos de visualización de datos. Por ejemplo, los algoritmos de auditoría pueden analizar los resultados de un modelo de IA para identificar cualquier disparidad en cómo se trata a diferentes grupos demográficos. Si se detectan sesgos, se puede ajustar el modelo o los datos para corregir estos problemas (Smith & Rustagi, 2021).

Además de las herramientas técnicas, es importante implementar procesos de revisión y auditoría continua para asegurar que los algoritmos se mantengan justos y equitativos a lo largo del tiempo.

Esto puede incluir la implementación de auditorías regulares de los algoritmos, así como la revisión de los datos de entrenamiento y los resultados del modelo para identificar y corregir cualquier sesgo que pueda surgir. La auditoría continua es esencial para asegurar que los algoritmos sigan siendo justos y equitativos a medida que se actualizan y evolucionan (Pannatier, 2022).

Las empresas tecnológicas también están comenzando a implementar herramientas de monitoreo y análisis en tiempo real para detectar y corregir sesgos en sus algoritmos. Estas herramientas pueden analizar los resultados de los algoritmos en tiempo real y alertar sobre cualquier problema de sesgo que pueda

surgir. Al detectar y corregir los sesgos en tiempo real, las empresas pueden asegurar que sus algoritmos sean justos y equitativos en todo momento (López, 2023).

Finalmente, la colaboración entre la industria y la academia es crucial para el desarrollo y la implementación de herramientas técnicas para detectar y corregir sesgos. Las investigaciones académicas pueden proporcionar nuevas ideas y enfoques para abordar el sesgo de género en la IA, mientras que la industria puede implementar y probar estas ideas en aplicaciones del mundo real. Al trabajar juntos, la industria y la academia pueden crear soluciones más efectivas y duraderas para el problema del sesgo de género en la IA (Lin, 2024).

4. Cambios necesarios para cerrar la Brecha de Género

Liderazgo intencional

El liderazgo intencional es fundamental para cerrar la brecha de género en la IA. Esto implica que el personal directivo de las empresas y entidades públicas no solo reconozca la importancia de la diversidad y la inclusión, sino que también se comprometa activamente a promover estos valores dentro de sus organizaciones. Lin (2024) enfatiza que el liderazgo intencional es crucial para asegurar que la equidad de género se convierta en una prioridad estratégica para las empresas tecnológicas.

Un aspecto clave del liderazgo intencional es la creación de políticas y prácticas que promuevan la inclusión y la diversidad. Esto puede incluir la implementación de programas de mentoría y desarrollo profesional para mujeres, así como la creación de oportunidades para que éstas asuman roles de liderazgo. Al apoyar activamente el desarrollo y la promoción de mujeres en la industria tecnológica, se puede ayudar a cerrar la brecha de género y crear un entorno más inclusivo y equitativo (Smith & Rustagi, 2021).

Además de las políticas y prácticas internas, quienes ostentan el liderazgo también pueden influir en el cambio a través de sus acciones y comportamientos. Al modelar un comportamiento inclusivo y apoyar activamente la diversidad, pueden inspirar a otras personas en la organización a hacer lo mismo. Esto puede incluir la participación en iniciativas de diversidad e inclusión, el apoyo a grupos de afinidad y la promoción de una cultura de respeto y equidad en el lugar de trabajo (Pannatier, 2022).

El liderazgo intencional también implica la rendición de cuentas. Se deben establecer métricas y objetivos claros para la diversidad y la inclusión, y ser responsables de alcanzar estos objetivos. Esto puede incluir la implementación de evaluaciones regulares del progreso hacia la igualdad de género, así cómo la creación de mecanismos para abordar cualquier problema o desafío que pueda surgir. Al ser transparentes y responsables en sus esfuerzos por promover la diversidad y la inclusión, quienes ostentan posiciones de liderazgo pueden asegurar que estos valores se integren de manera efectiva en la cultura y las operaciones de la organización (Lin, 2024).

Finalmente, el liderazgo intencional también debe incluir la colaboración y la asociación con otras organizaciones y grupos de defensa de la igualdad de género. Al trabajar juntos, las empresas y las organizaciones de defensa pueden compartir recursos, conocimientos y mejores prácticas, lo que puede ayudar a crear un impacto más significativo y duradero. Las alianzas estratégicas pueden proporcionar el apoyo y la experiencia necesarios para abordar los desafíos de laequidad de género en la IA y promover un cambio positivo en toda la industria (Wisner Glusko, 2022).

Alianzas estratégicas

Las alianzas estratégicas son esenciales para cerrar la brecha de género en la IA. Estas alianzas pueden incluir colaboraciones entre empresas, organizaciones no gubernamentales, instituciones académicas y grupos de defensa de la igualdad. Al trabajar juntos, estas entidades pueden compartir recursos, conocimientos y mejores prácticas.

Una de las formas más efectivas de alianza estratégica es la creación de consorcios y coaliciones que se centren en promover la diversidad y la inclusión en la IA.

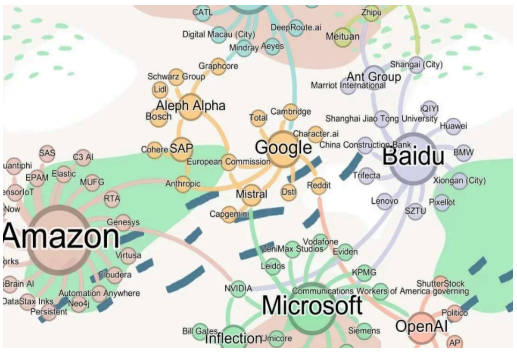
Estos consorcios pueden reunir a diversas partes interesadas para trabajar en proyectos conjuntos, desarrollar estándares y directrices, y abogar por políticas que promuevan la igualdad en la IA. Al unirse en torno a un objetivo común, estas alianzas pueden amplificar sus esfuerzos y lograr un cambio más rápido y efectivo (López, 2023).

Las alianzas estratégicas también pueden incluir la colaboración en proyectos de investigación y desarrollo. Las empresas y las instituciones académicas pueden trabajar juntas para investigar y desarrollar nuevas tecnologías y enfoques que aborden el sesgo de género en la IA. Estas colaboraciones pueden proporcionar a las empresas acceso a los últimos avances científicos y técnicos, mientras que las instituciones académicas pueden beneficiarse de la experiencia práctica y los recursos de las empresas.

Otra forma importante de alianza estratégica es la colaboración en la educación y la formación. Las empresas pueden trabajar con instituciones educativas para desarrollar programas de formación y certificación en temas de equidad y ética en la IA. Estos programas pueden equipar a quienes dirigirán el futuro de la IA con las habilidades y el conocimiento necesarios para crear tecnologías más inclusivas y equitativas. Al apoyar la educación y la formación en estos temas, las empresas pueden ayudar a preparar a la próxima generación de desarrolladores y desarrolladoras de IA para abordar sesgos de manera efectiva (Sentance, 2022).

Finalmente, las alianzas estratégicas también pueden incluir la colaboración en la promoción y la defensa de políticas que promuevan la equidad de género en la IA. Las empresas y las organizaciones de defensa pueden trabajar juntas para abogar por regulaciones y políticas que aseguren que los datos y los algoritmos de IA sean justos y

equitativos. Al unirse en torno a estas causas, estas alianzas pueden ejercer una mayor influencia y lograr un cambio más significativo en la industria y en la sociedad en general (Wisner Glusko, 2022).



Mapa de las Alianzas IA. La clave para tejer la red del futuro en inteligencia artificial

¿Por qué las colaboraciones estratégicas son importantes en la adopción de la IA?

Toma de riesgos

La toma de riesgos es un aspecto crucial para cerrar la brecha en la IA. Las mujeres deben ser alentadas y apoyadas para asumir roles de liderazgo y tomar riesgos en el desarrollo y la implementación de tecnologías de IA. Esto es particularmente importante en áreas impactadas por la IA generativa, donde la innovación y el liderazgo pueden tener un impacto significativo en la dirección futura de la tecnología (Lin, 2024).

Una de las formas en que las empresas pueden fomentar la toma de riesgos es a través de programas de mentoría y desarrollo profesional. Estos programas pueden proporcionar a las mujeres el apoyo y la orientación necesarios para asumir roles de liderazgo y tomar riesgos calculados en sus carreras. Al empoderar a las mujeres para que se conviertan en líderes, las empresas pueden ayudar a cerrar la brecha de género y promover una mayor diversidad en la industria tecnológica (Smith & Rustagi, 2021).

Además de los programas de mentoría, las empresas también pueden crear entornos de trabajo que fomenten la toma de riesgos y la innovación. Esto puede incluir la implementación de políticas y prácticas que promuevan la creatividad y la experimentación, así como la creación de una cultura que valore y celebre el fracaso como una oportunidad de aprendizaje. Al proporcionar un entorno seguro y de apoyo para la toma de riesgos, las empresas pueden animar a las mujeres a asumir desafíos y explorar nuevas ideas y enfoques (López, 2023).

La toma de riesgos también implica la disposición a desafiar el status quo y abogar por el cambio. Las mujeres en roles de liderazgo pueden desempeñar un papel crucial en la promoción de políticas y prácticas más inclusivas y equitativas en la IA. Esto puede incluir abogar por la diversificación de los datos de entrenamiento, la implementación de herramientas técnicas para detectar y corregir sesgos, y la promoción de la educación y la conciencia sobre el sesgo de género en la IA.

Finalmente, la toma de riesgos también puede incluir la colaboración y la asociación con otras organizaciones y grupos de defensa por la igualdad. Al trabajar juntos, las empresas y las organizaciones de defensa pueden compartir recursos, conocimientos y mejores prácticas, lo que puede ayudar a crear un impacto más significativo y duradero en esta lucha (Wisner Glusko, 2022).

La brecha de género en la era de la Inteligencia Artificial

Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España

La brecha de género también se extiende a la IA

Políticas y regulaciones

Una de las áreas clave donde las políticas pueden tener un impacto significativo es en la diversificación de los datos de entrenamiento. Las políticas pueden requerir que las empresas utilicen conjuntos de datos diversos y representativos para entrenar sus algoritmos, asegurando que los datos reflejen una variedad de contextos y experiencias. Esto puede ayudar a mitigar los sesgos en los algoritmos y a promover una mayor equidad en los resultados de la IA (López, 2023).

Además de la diversificación de datos, las políticas también pueden establecer requisitos para la revisión y auditoría continua de los algoritmos de IA. Las auditorías regulares pueden ayudar a identificar y corregir cualquier sesgo que pueda surgir en los algoritmos, asegurando que estos se mantengan justos y equitativos a lo largo del tiempo. Las políticas también pueden exigir que las empresas implementen herramientas de monitoreo y análisis en tiempo real para detectar y corregir sesgos en sus algoritmos (Pannatier, 2022).

Las políticas y regulaciones también pueden fomentar la educación y la formación en temas de equidad y ética en la IA. Esto puede incluir la creación de programas de formación y certificación que equipen con las habilidades y el conocimiento necesarios para abordar el sesgo de manera efectiva.

Finalmente, las políticas y regulaciones pueden promover la inclusión de perspectivas de género en los equipos de desarrollo de IA. Esto puede incluir incentivos para la contratación y promoción de mujeres en roles técnicos y de liderazgo, así como la creación de programas de mentoría y desarrollo profesional para mujeres.

Al fomentar la diversidad de género en los equipos de desarrollo, las políticas pueden ayudar a crear un entorno más inclusivo y equitativo para el desarrollo de la IA (Smith & Rustagi, 2021).

Observatorio de Innovación Educativa y Cultura Digital

Desarrollo inclusivo y diversidad en la inteligencia artificial

¿Cómo utilizar la inteligencia artificial para promover la diversidad en el lugar de trabajo?

Inteligencia artificial y equidad de género: un espejo de nuestras sociedades

Inteligencia artificial con perspectiva de género

Explorando la Intersección de Género, Tecnología e IA

Evaluaciones de impacto y transparencia

Las evaluaciones de impacto y la transparencia son cruciales. Las evaluaciones de impacto pueden ayudar a identificar y mitigar los efectos negativos de los algoritmos de IA en diferentes grupos demográficos, asegurando que los algoritmos sean justos y equitativos.

La transparencia, por otro lado, permite que los desarrolladores, los usuarios y el público en general comprendan cómo funcionan los algoritmos y cómo se toman las

decisiones, lo que es esencial para fomentar la confianza y la rendición de cuentas (Pannatier, 2022).

Una de las formas en que las evaluaciones de impacto pueden ser implementadas es a través de auditorías externas e independientes. Estas auditorías pueden revisar los algoritmos y los datos de entrenamiento para identificar cualquier sesgo y recomendar cambios para corregir estos problemas. Las auditorías externas pueden proporcionar una perspectiva objetiva y neutral, lo que puede ayudar a asegurar que los algoritmos sean justos y equitativos (López, 2023).

Además de las auditorías externas, las empresas también pueden implementar procesos internos de revisión y auditoría. Estos procesos pueden incluir la revisión regular de los algoritmos y los datos de entrenamiento, así como la implementación de herramientas técnicas para detectar y corregir sesgos. Al establecer procesos internos de revisión y auditoría, las empresas pueden asegurarse de que sus algoritmos se mantengan justos y equitativos a lo largo del tiempo (Smith & Rustagi, 2021).

La transparencia es otro aspecto crucial para cerrar la brecha de género en la IA. Las empresas deben ser transparentes sobre cómo funcionan sus algoritmos y cómo se toman las decisiones. Esto puede incluir la publicación de información sobre los datos de entrenamiento, los modelos de algoritmos y los resultados de las evaluaciones de impacto.

La transparencia permite que los desarrolladores, los usuarios y el público en general comprendan cómo funcionan los algoritmos y cómo se toman las decisiones, lo que es esencial para fomentar la confianza y la rendición de cuentas (Lin, 2024).

Finalmente, la transparencia también puede incluir la creación de mecanismos para que los usuarios y usuarias proporcionen retroalimentación y denuncien problemas. Esto puede incluir la implementación de sistemas de retroalimentación en línea, la creación de comités de ética y la promoción de la participación del públco en el desarrollo y la revisión de los algoritmos. Al fomentar la participación y la retroalimentación, las empresas pueden identificar y abordar problemas de manera más efectiva, asegurando que sus algoritmos sean justos y equitativos (Wisner Glusko, 2022).

Inteligencia artificial: Transparencia

Evaluaciones de impacto de derechos fundamentales sobre sistemas de IA de alto riesgo en el RIA

Transparencia en el uso de algoritmos de inteligencia artificial

Campañas de sensibilización y educación

Estas campañas pueden destacar la importancia de la diversidad y la inclusión en la IA, y proporcionar información sobre cómo el sesgo de género puede afectar a los algoritmos y a las decisiones automatizadas. Al sensibilizar al público y a la industria sobre estos problemas, las campañas de sensibilización pueden ayudar a fomentar un cambio cultural en la industria tecnológica (Sentance, 2022).

La promoción y defensa de la igualdad de género también puede incluir la participación en conferencias y eventos de la industria. Estos eventos pueden proporcionar una plataforma para discutir y abordar los problemas de equidad de género en la IA, y para compartir mejores prácticas y soluciones innovadoras. Al participar en estos eventos, las empresas y las organizaciones pueden ayudar a fomentar un diálogo constructivo y promover un cambio positivo en la industria tecnológica (Wisner Glusko, 2022).

Inteligencia artificial, un nuevo horizonte en la lucha contra la violencia de género

CAMPAÑA #ROMPEELSESGODIGITALFM EL IMPACTO DE LAS TIC EN LOS PROCESOS DE SELECCIÓN: SESGOS DE GÉNERO EN LA IA

La inteligencia artificial en la educación

Recomendación sobre la ética de la inteligencia artificial

Ética de la inteligencia artificial

Bibliografía

Doshi, T. (2018). How Google's Inclusive Images Project is Diversifying Data for AI. *Google AI Blog.* Recuperado de https://ai.googleblog.com/2018/09/how-googles-inclusive-images-project-is.html

Lin, S. (2024). Closing the Gender Gap in AI: Best Practices for Inclusivity. IBM Blog. Recuperado de https://www.ibm.com/blog/closing-the-gender-gap-in-ai/

Lin, S. (2024). The Role of Leadership in Addressing Gender Bias in AI. TechEthics Journal. Recuperado de https://www.techethicsjournal.com/leadership-gender-bias-ai

López, A. (2023). Gender Bias in AI: An Urgent Need for Diverse Perspectives. *TechCrunch.* Recuperado de https://techcrunch.com/2023/04/10/gender-bias-in-ai

López, A. (2023). Collaborative Efforts to Mitigate Gender Bias in AI Development. AI Policy Forum. Recuperado de https://www.aipolicyforum.org/mitigating-gender-bias

Pannatier, E. (2022). Continuous Auditing in AI: Ensuring Fairness and Equity. *AI Ethics Review.* Recuperado de https://www.aier.org/continuous-auditing-in-ai

Pannatier, E. (2022). Addressing Gender Bias in AI: Challenges and Opportunities. *Le Temps.* Recuperado de https://www.letemps.ch/sciences/addressing-gender-bias-ai

Sentance, S. (2022). Gender Bias in AI: Tackling the Problem Head-On. *Raspberry Pi Computing Education Research Centre.* Recuperado de https://www.raspberrypi.org/blog/gender-bias-in-ai/

Smith, G., & Rustagi, I. (2021). Discrimination in the age of algorithms. *Journal of Legal Analysis*, 11(1), 63-115. DOI: 10.1093/jla/laaa005.

Wisner Glusko, D. C. (2022). Legal Frameworks for Mitigating Gender Bias in AI. *European Journal of Law and Technology.* Recuperado de https://www.ejlt.org/article/view/856

Wisner Glusko, D. C. (2022). Strategic Alliances to Address Gender Bias in AI. *Gender and Technology Insights.* Recuperado de https://www.gendertechinsights.com/strategic-alliances-ai

Tema 6. Conclusiones y recomendaciones

Presentación

En la última unidad del curso se abordan las conclusiones y recomendaciones para superar los desafíos del sesgo de género en la inteligencia artificial (IA). A lo largo de la formación se ha explorado cómo la IA puede perpetuar desigualdades de género debido a la falta de representación en datos de entrenamiento, decisiones de diseño, y falta de diversidad en los equipos de desarrollo.

Esta unidad recopila los hallazgos clave de las unidades anteriores y proporciona recomendaciones prácticas para construir una IA más inclusiva y equitativa. La integración de la igualdad de género en el desarrollo y aplicación de la IA es crucial para asegurar que la tecnología beneficie a toda la sociedad y no exacerbe las desigualdades ya existentes.

Objetivos

Los objetivos fundamentales que la alumna podrá alcanzar una vez terminado el estudio de este tema son los siguientes:

- **Resumir los hallazgos sobre el sesgo de género en la IA**: identificar cómo la IA puede perpetuar o amplificar las desigualdades de género y las áreas más afectadas por estos sesgos.
- **Proponer soluciones prácticas para mitigar el sesgo de género en la IA**: desarrollar estrategias para diversificar datos de entrenamiento, implementar mejores prácticas en el diseño de algoritmos y fomentar la diversidad en equipos de desarrollo.
- **Desarrollar recomendaciones prácticas para la equidad de género en IA**: proporcionar directrices para empresas y organizaciones sobre cómo abordar el sesgo de género en sus prácticas de IA.
- **Identificar las mejores prácticas para la inclusión de género en IA**: examinar casos exitosos de inclusión de género en proyectos de IA y extraer lecciones aplicables de cada uno de ellos.

1. Conclusiones y recomendaciones

El sesgo de género en la inteligencia artificial es un problema complejo y multifacético que requiere un enfoque integral para ser abordado, tal y como hemos ido viendo a lo largo del curso. Desde la subrepresentación de mujeres en roles técnicos hasta los sesgos inherentes en los datos de entrenamiento, las causas del sesgo de género en la IA son diversas. Cerrar la brecha de género en la IA pasa por

diversas acciones, incluyendo la diversificación de los equipos de desarrollo, la implementación de herramientas técnicas para detectar y corregir sesgos, y la creación de un marco regulatorio robusto.

La diversidad en los equipos de desarrollo es crucial para asegurar que se consideren diversas perspectivas y experiencias en el diseño y la implementación de algoritmos. La diversificación de los datos de entrenamiento también resulta clave para mitigar los sesgos en los algoritmos y promover una mayor equidad en los resultados de la IA. Además, la implementación de herramientas técnicas para detectar y corregir sesgos puede ayudar a asegurar que los algoritmos sean justos y equitativos (López, 2023; Pannatier, 2022).

Conseguir una IA que no olvide la ética es uno de los grandes desafíos que plantea esta tecnología

IA y Ética: Navegando en los desafíos de la era tecnológica

Inteligencia Artificial y Ética (II): El desafío del sesgo

El liderazgo intencional y la toma de riesgos son fundamentales para promover la igualdad de género en la IA. Las personas que toman decisiones en la industria deben comprometerse activamente a promover la diversidad y la inclusión, y responsabilizarse de alcanzar estos objetivos. La promoción y defensa de la igualdad de género también son esenciales para sensibilizar al público sobre la importancia de estos valores, y a la hora de abogar por políticas y prácticas más inclusivas y equitativas (Lin, 2024 ; Wisner Glusko, 2022).

Finalmente, las alianzas estratégicas y la colaboración entre empresas, organizaciones, instituciones académicas y otras partes interesadas pueden desempeñar un papel crucial en la promoción de la igualdad de género en la IA. Estas alianzas pueden proporcionar recursos, conocimientos y mejores prácticas, y pueden ayudar a crear un impacto más significativo y duradero en la lucha contra el sesgo de género en la IA (Smith & Rustagi, 2021; López, 2023).

En definitiva, abordar el sesgo en la IA requiere considerar aspectos éticos, sociales y técnicos que impactan en el diseño, implementación y uso de sistemas basados en IA (UNESCO, 2024).

Facilitar la inclusión de la perspectiva de sexo y género en políticas públicas, programas educativos, la industria y la investigación es clave para ayudar a superar los sesgos de género de la IA en sectores tan esenciales como la salud o la educación, por citar solo dos de los ámbitos donde dicha inteligencia ya nos impacta directamente.

Líneas de acción para superar los sesgos de sexo y género en ciencia |

Fundación La Caixa https://www.youtube.com/watch?v=DrdfS3dSNf4

Fundación La Caixa
https://www.youtube.com/watch?v=RtQyWxzAnVY&list=PLp5NbUWNrI9exC8Ivprp0DTRFhmb6oNoJ&index=409

Pero no basta únicamente con incluir la perspectiva de género en la IA: hay que aumentar la representación de las mujeres en la ciencia, tecnología, ingeniería y matemáticas (las denominadas carreras STEM), para combatir el sesgo de sexo y género existente y emergente en estos campos (UNESCO, 2024).

Cuando la Comisión Europea publicó su esperado libro blanco *Sobre la Inteligencia Artificial - Un enfoque europeo hacia la excelencia y la confianza*, en 2020, gran parte de la reacción inicial del público se centró en la potencial regulación de la IA, lo que supone un desafío adicional para la posición de la UE ante la feroz competencia tecnológica de China y Estados Unidos. Pocos discutieron la mención del documento de la Comisión Europea sobre las directrices éticas y de género.

El libro blanco solicita "requisitos para tomar medidas razonables destinadas a garantizar que el uso de los sistemas de IA no conduzca a resultados que impliquen discriminación ". ¿Por qué es esto importante?

Este enfoque no es solo teórico respecto a la discriminación. También se trata en gran medida de salvar vidas, especialmente de mujeres, y asegurar que los productos y servicios esenciales satisfagan las necesidades tanto de mujeres como de hombres. Y es que, si la inteligencia artificial se basa en datos "malos", predominantemente de hombres o basados en perfiles masculinos, pueden ocurrir cosas terribles si nos fijamos únicamente, por citar un caso, en el ámbito de la salud. Por ejemplo, cinturones de seguridad, reposacabezas y bolsas de aire en automóviles que han sido diseñados principalmente con datos obtenidos de pruebas de choque utilizando la fisonomía masculina. Los cuerpos de mujeres embarazadas y sus características no se incorporan en las "mediciones estándar". Cómo resultado, las mujeres tienen un 47% más de probabilidades de sufrir lesiones graves y un 17% más de morir en un accidente similar que la de un hombre, según explican Caroline Criado Perez, autora de "Invisible Women", y Lauren Klein, coautora de "Data Feminism", en una reciente entrevista con la BBC.

¿Cómo la IA puede ser una herramienta para favorecer la igualdad de género ?

La inteligencia artificial como herramienta para la equidad de género

Y en este mismo año 2024, la nueva ley de Inteligencia Artificial ha reafirmado el compromiso de la Comisión Europea para garantizar que los sistemas de Inteligencia Artificial utilizados en la Unión Europea respeten los derechos de la ciudadanía.

Obtén más información sobre esta ley clave

Las claves de la nueva ley de Inteligencia Artificial
¿Sabías que la UE será la primera en establecer reglas claras para el uso de la Inteligencia Artificial? Te contamos en qué van a consistir.

2. Recomendaciones prácticas

Basado en todo lo planteado hasta ahora en el curso, se pueden hacer varias recomendaciones prácticas para incrementar la igualdad de género en los sistemas y productos derivados de la IA.

En primer lugar, las empresas deben comprometerse a diversificar sus equipos de desarrollo, implementando políticas y prácticas que promuevan la contratación y la promoción de mujeres en roles técnicos y de liderazgo. Esto puede incluir la creación de programas de mentoría y desarrollo profesional, así como la implementación de incentivos para la contratación y la promoción de mujeres (Smith & Rustagi, 2021; Lin, 2024).

Segundo, las empresas deben diversificar sus datos de entrenamiento para asegurar que los algoritmos sean justos y equitativos. Esto puede incluir la recolección de nuevos datos más representativos, la eliminación de datos sesgados y el ajuste de los algoritmos para mitigar los efectos del sesgo. Las auditorías regulares y las herramientas de monitoreo en tiempo real también son esenciales para asegurar que los algoritmos se mantengan justos y equitativos a lo largo del tiempo (López, 2023; Pannatier, 2022).

¿Cómo garantizar algoritmos libres de sesgos de género ?

Garantizar La Equidad Y Los Algoritmos Libres De Sesgos En La Tecnología Educativa

Garantizar la equidad y la transparencia

Tercero, las empresas deben ser transparentes sobre cómo funcionan sus algoritmos y cómo se toman las decisiones. Esto puede incluir la publicación de información sobre los datos de entrenamiento, los modelos de algoritmos y los resultados de las evaluaciones de impacto. La transparencia permite que quienes desarrollan y utilizan los productos resultantes comprendan cómo funcionan los algoritmos y cómo se toman las decisiones, lo que es esencial para fomentar la confianza y la rendición de cuentas (Lin, 2024; Wisner Glusko, 2022).

Cuarto, las empresas y organizaciones deben apoyar la educación y la formación en temas de igualdad de género y ética en la IA. Esto puede incluir la creación de programas de formación y certificación que equipen a quienes desarrollan y lideran la industria con las habilidades y el conocimiento necesarios para abordar el sesgo de género de manera efectiva. Al apoyar la educación y la formación en estos temas, las empresas pueden ayudar a preparar a la próxima generación de desarrolladores y desarrolladoras de IA para crear tecnologías más inclusivas y equitativas (Sentance, 2022; López, 2023).

Finalmente, las empresas deben colaborar con otras organizaciones y grupos de defensa de la igualdad de género para promover la diversidad y la inclusión en la IA. Las alianzas estratégicas pueden proporcionar recursos, conocimientos y mejores

prácticas, y pueden ayudar a crear un impacto más significativo y duradero en la lucha contra el sesgo de género en la IA. Al trabajar juntos, las empresas y las organizaciones de defensa pueden compartir recursos, conocimientos y mejores prácticas, lo que puede ayudar a crear un impacto más significativo y duradero en la promoción de la equidad de género en la IA (Wisner Glusko, 2022; Smith & Rustagi, 2021).

Cómo diseñamos el futuro de la IA dice mucho de nuestro avance como sociedad:

Inteligencia artificial y equidad de género: un espejo de nuestras sociedades

En definitiva, cerrar la brecha de género en la inteligencia artificial es un desafío complejo que requiere un enfoque integral y multifacético. Desde la diversificación de los equipos de desarrollo y los datos de entrenamiento hasta la implementación de herramientas técnicas y la creación de un marco regulatorio robusto, existen muchas estrategias que pueden ayudar a abordar los sesgos en la IA. Al promover la diversidad y la inclusión, la educación y la formación, y la colaboración y la alianza estratégica, las empresas y las organizaciones pueden crear un impacto más significativo y duradero en la lucha contra estos (Smith & Rustagi, 2021; López, 2023; Lin, 2024).

El liderazgo intencional y la toma de riesgos son fundamentales para promover las equidades en la IA. La promoción y defensa de la equidad también son esenciales para sensibilizar al público y a la industria sobre la importancia de estos valores, y para abogar por políticas y prácticas más inclusivas y equitativas (Lin, 2024; Wisner Glusko, 2022).

Finalmente, las investigaciones futuras deben centrarse en desarrollar nuevas técnicas y enfoques para detectar y corregir los sesgos en la IA, y en explorar cómo las políticas, la educación y la colaboración pueden promover la equidad en la IA. Al trabajar de manera conjunta, la industria, la academia y las organizaciones de defensa pueden crear soluciones más innovadoras y efectivas para el problema del sesgo de género en la IA (Smith & Rustagi, 2021; López, 2023; Lin, 2024).

Bibliografía

Aguerre, M., Balmaceda, N., López, M., Peller, F., Tagliazucchi, G., & Zeller, F. (2023). Reconsidering AI: Ethical Challenges and Opportunities. *TechEthics Journal.* Recuperado de https://www.techethicsjournal.com/reconsidering-ai

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, 77-91. Recuperado de https://proceedings.mlr.press/v81/buolamwini18a.html

Captain, S. (2015). Amazon Killed AI Hiring Tool That Showed Bias Against Women. *TechCrunch.* Recuperado de https://techcrunch.com/2015/10/09/amazon-killed-ai-hiring-tool

Dastin, J. (2018). AI Can Be Sexist Too: How Artificial Intelligence Is Replicating the Biases of Its Creators. *Reuters.* Recuperado de https://www.reuters.com/article/us-amazon-com-employment-idUSKCN1MK0O3

González Peña, A. (2023). The Role of AI in Modern Sexual Education. *Educational Technology & Society.*, 26(2), 212-223. Recuperado de https://www.j-ets.net/ETS/journals/26_2/22.pdf

Gray, M. L. (2021). Bias in AI: A Queer Perspective on Algorithmic Discrimination. *Journal of Technology and Society.* Recuperado de https://www.jtechsoc.org/article/view/1234

Johnson, K. (2023). AI and Gender Bias: Understanding the Impact. *Harvard Business Review.* Recuperado de https://hbr.org/2023/02/ai-and-gender-bias

Mohan, S. (2023). AI generativa y sesgos de género : Un desafío persistente. *Fronteras Digitales del Conocimiento.* Recuperado de https://www.orfonline.org/expert-speak/gender-ative-ai

Nicoletti, L., & Bass, D. (2023). Humans Are Biased. Generative AI Is Even Worse. *TechCrunch.* Recuperado de https://techcrunch.com/2023/06/09/humans-are-biased-generative-ai-is-even-worse

Niethammer, C. (2020). The Tech Gender Gap: Breaking Barriers in the Age of AI. *The Economist.* Recuperado de https://www.economist.com/special-report/2020/10/17/the-tech-gender-gap-breaking-barriers

Sandoval-Martin, J., & Martínez-Sanzo, I. (2024). Gender Representation in AI-Generated Images: A Comparative Study. *Digital Arts and Ethics Journal.* Recuperado de https://www.daejournal.com/gender-representation-in-ai-generated-images

UNESCO. (2024). Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes. Recuperado de https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes.

UN Women. (2023). AI perpetuates stereotypes: Analysis of image generation models. Recuperado de https://www.unwomen.org/en/news-stories/feature-story/2023/03/ai-perpetuates-stereotypes-analysis-of-image-generation-models

Vázquez Figueiredo, C., Pascual, R. F., & González Peña, A. (2023). AI's Influence on Sexual Education Among Adolescents: A Double-Edged Sword.
The Conversation.
Recuperado de https://theconversation.com/ai-influence-on-sexual-education-among-adolescents-205312

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

VICERRECTORADO DE ARTE, CIENCIA, TECNOLOGÍA Y SOCIEDAD

GENERALITAT VALENCIANA

Vicepresidencia Segunda y Conselleria de Servicios Sociales, Igualdad y Vivienda

Algoritmos con perspectiva de género.

Cómo la IA puede eliminar sesgos discriminatorios sobre la mujer.