

Inteligencia Artificial con perspectiva de género

**Guía de buenas prácticas para detectar, prevenir
y mitigar los sesgos de género en el desarrollo y
uso de la IA**



Guía de buenas prácticas para detectar, prevenir y mitigar los sesgos de género en el desarrollo y uso de la IA

Departamento Nacional de Planeación (DNP)

Natalia Irene Molina Posso

Directora General DNP

Rovitzon Ortiz Olaya

Subdirector General de Prospectiva y Desarrollo Nacional

Edwin Alejandro Buenhombre Moreno

Director de Desarrollo Digital

Equipo autor

Viviana Vanegas Barrero
Yenifer Julie Pinto Gaitán
Sara Daniela Márquez Gutiérrez
Natalia Andrea Ramírez Pérez

Este documento contiene insumos técnicos y metodológicos desarrollados por



Resumen

La Guía de buenas prácticas para detectar, prevenir y mitigar los sesgos de género en el desarrollo y uso de la inteligencia artificial (IA) es una iniciativa del Departamento Nacional de Planeación (DNP) con la cual se busca enfrentar el riesgo de que la inteligencia artificial (IA) replique y amplifique las desigualdades de género. Es un documento práctico que ofrece un marco de acción para el ecosistema digital colombiano, con orientaciones y recursos que pueden usar todas las personas y organizaciones que participan en el ciclo de vida de los sistemas de IA, desde su diseño hasta su adopción.

La guía incluye dos instrumentos prácticos un Catálogo de Herramientas y un Árbol de Decisión. Su objetivo es proporcionar los medios para detectar, prevenir y mitigar sesgos de género, asegurando que la innovación tecnológica contribuya a la equidad, la justicia social y el desarrollo de una IA confiable, justa y responsable en Colombia.

Contenido

Conceptos clave para navegar esta guía	6
1. ¿Por qué necesitamos una guía para prevenir los sesgos de género en sistemas de inteligencia artificial?	10
El compromiso de Colombia por impulsar una transformación digital con equidad de género	12
Diálogos globales: marcos internacionales para una IA ética e inclusiva	14
Referentes semejantes a esta guía de buenas prácticas	14
¿A quiénes convoca esta guía?	16
¿Qué contiene esta guía?	17
3. ¿Cómo se reproducen desigualdades de género en sistemas de inteligencia artificial?.....	19
¿Cuáles son los sesgos de género en sistemas de IA?.....	20
4. ¿Cómo podemos promover la equidad de género en los sistemas de inteligencia artificial?	24
Buenas prácticas para el desarrollo de sistemas de IA.....	24
Buenas prácticas para la adopción de los sistemas de IA.....	33
5. Árbol de decisión y catálogo de herramientas para la detección, prevención y mitigación de sesgos de género en sistemas de IA.....	35
Árbol de decisión	35
¿Cómo se hizo?	36
¿Para qué sirve?.....	37
¿Cómo se usa?	38
Catálogo de herramientas	41
¿Qué es?	41
¿Para qué sirve?.....	42
¿Cómo se usa?	42
Referencias	45

Índice de Figuras

<i>Figura 1 Guía para una IA equitativa en Colombia</i>	<i>18</i>
<i>Figura 2 Ruta de navegación del árbol de decisión</i>	<i>38</i>
<i>Figura 3 Representación gráfica del árbol de decisión: navegación guiada e interactiva</i>	<i>40</i>
<i>Figura 4 Catálogo de Herramientas: estructura y filtros para explorar recursos</i>	<i>44</i>

Índice de Tablas

<i>Tabla 1 Principales sesgos en sistemas de inteligencia artificial</i>	<i>21</i>
<i>Tabla 2 Buenas prácticas en la etapa de planificación y diseño</i>	<i>24</i>
<i>Tabla 3 Buenas prácticas en la etapa de recopilación y tratamiento de datos</i>	<i>26</i>
<i>Tabla 4 Buenas prácticas en la etapa de construcción y adaptación de modelos</i>	<i>28</i>
<i>Tabla 5 Buenas prácticas en la etapa de prueba, evaluación, verificación y validación</i>	<i>29</i>
<i>Tabla 6 Buenas prácticas en la etapa de entrada en servicio o despliegue</i>	<i>31</i>
<i>Tabla 7 Buenas prácticas en la etapa de explotación y supervisión (Monitoreo y ajuste)</i>	<i>31</i>
<i>Tabla 8 Buenas prácticas en la etapa de retirada/desmantelamiento</i>	<i>32</i>
<i>Tabla 9 Buenas prácticas para la adopción de los sistemas de IA</i>	<i>33</i>

Conceptos clave para navegar esta guía

Considerando la importancia de la inteligencia artificial con perspectiva de género, a continuación se presenta un glosario de términos clave, derivado de los avances técnicos, académicos y de gestión que permiten tener el contexto marco de la presente: “Guía de buenas prácticas para detectar, prevenir y mitigar los sesgos de género en el desarrollo y uso de la IA”.

- ✧ **Inteligencia artificial (IA):** se refiere a la capacidad de las máquinas o sistemas informáticos para realizar tareas que normalmente requieren inteligencia humana, como el reconocimiento de patrones, la toma de decisiones, la resolución de problemas y el aprendizaje a partir de la experiencia. Los sistemas de IA utilizan diversas técnicas, como el aprendizaje automático, el procesamiento de lenguaje natural y las redes neuronales, para realizar estas tareas de manera eficiente y adaptativa (Stuart J & Norvig, 2021; Angwin, Larson, Mattu, & Kirchner, 2016).
- ✧ **Ciclo de vida de la IA:** se refiere a las distintas fases que comprende el diseño de sistemas de IA: planificación y diseño; recopilación y tratamiento de datos; creación de modelo(s) y/o adaptación de modelo(s) existente(s) a tareas específicas; prueba, evaluación, verificación y validación; entrada en servicio/despliegue; explotación y supervisión; y retirada/desmantelamiento. Estas etapas no son secuenciales, sino iterativas (OCDE, 2024).
- ✧ **Gobernanza ética de IA:** se refiere al conjunto de marcos, principios, políticas, estándares y mecanismos institucionales que orientan el diseño, desarrollo, despliegue y uso de los sistemas de IA de manera ética, transparente y responsable. Su propósito es anticipar y mitigar riesgos potenciales, incluidos sesgos y desigualdades, para asegurar que estas tecnologías se alinean con los derechos humanos, la equidad y el interés público.
- ✧ **Datos:** son el insumo base para el entrenamiento de los sistemas de IA, pueden ser texto, imágenes, audio, video, registros numéricos, entre otros. Si los datos no representan de forma equitativa a mujeres, personas LGBTIQ+ u otras identidades de género, el sistema aprenderá con una mirada parcializada.
- ✧ **Código:** es la forma en que los algoritmos y modelos son escritos y ejecutados. El lenguaje de programación no tiene género, pero quienes diseñan, prueban y validan el código sí toman decisiones que afectan qué se prioriza y cómo se implementa.
- ✧ **Algoritmos:** son las instrucciones o conjunto ordenado de pasos que permiten a la IA tomar decisiones. Aunque los algoritmos en sí no “piensan”, replican patrones presentes en los datos con los que fueron entrenados. Por eso, si los datos tienen sesgos, las decisiones del sistema también los tendrán.

- ☼ **Sesgo algorítmico:** decisiones sistemáticamente desfavorables o favorables hacia ciertos grupos o individuos debido a fallas o características inherentes al diseño o los datos utilizados por un algoritmo (Business & Marketing School, 2024).
- ☼ **Transparencia algorítmica:** se refiere a la capacidad de los sistemas basados en algoritmos para ser comprensibles, explicables y auditables por los usuarios y reguladores. Implica que las decisiones tomadas por algoritmos, especialmente en áreas críticas como el acceso a servicios, la justicia o la contratación, deben ser claras en cuanto a los criterios utilizados para llegar a esas decisiones (Kossow, Svea, Svea , & Jenkins, 2021).
- ☼ **Explicabilidad:** se refiere a la capacidad de interpretar y justificar las decisiones detrás del diseño de un modelo o sistemas de IA. Promover la explicabilidad es fundamental para garantizar que los sistemas de IA sean auditables y confiables para todas las personas, especialmente aquellas que históricamente han sido excluidas.
- ☼ **Auditoría algorítmica:** evaluación sistemática de los algoritmos de IA para identificar, medir y corregir sesgos o errores que puedan generar decisiones discriminatorias. Las auditorías algorítmicas son esenciales para garantizar la transparencia y equidad en los sistemas de IA (Raji & Buolamwini, 2019).
- ☼ **Equidad algorítmica:** es tanto un principio ético para el diseño, desarrollo y uso de algoritmos, con el cual se busca garantizar que estos operen de manera justa, equitativa, imparciales y no discriminatoria, evitando la reproducción o amplificación de sesgos sistemáticos.
- ☼ **Modelo:** es la representación final que aprende a partir de los datos y los algoritmos. El modelo, por ejemplo, recomienda, clasifica o predice algo. Si no es supervisado con perspectiva de género, puede tomar decisiones inequitativas sin que quienes lo usan se den cuenta.
- ☼ **Entrenamiento de modelos:** proceso iterativo mediante el cual un modelo de aprendizaje automático aprende pautas y relaciones a partir de datos de ejemplo, ajustando sus parámetros internos para mejorar su capacidad de realizar predicciones o tomar decisiones (Goodfellow, Bengio, & Courville, 2016). Durante esta etapa, el algoritmo analiza patrones y relaciones en los datos para generar un modelo que pueda predecir, clasificar o recomendar algo en el futuro. Si los datos usados para entrenar el modelo contienen desigualdades, omisiones o estereotipos de género, el modelo las puede replicar o amplificar. Por eso, es crucial que esta etapa incluya criterios de equidad y validación diversa.
- ☼ **Mantenimiento de un modelo de IA:** proceso mediante el cual se monitorea el desempeño de un sistema de IA, se identifican errores o sesgos emergentes, y se hacen

los ajustes necesarios según los hallazgos. El mantenimiento no es solo técnico: también debe incorporar criterios éticos y de equidad, revisando si los datos siguen siendo representativos o si el modelo toma decisiones que afectan la equidad de género.

- ✧ **Género:** conjunto de características socialmente construidas que definen los roles, comportamientos y relaciones entre mujeres, hombres y personas sexualmente diversas. Estas características varían según el contexto cultural, social, económico y temporal, y suelen influir en las oportunidades y los derechos de las personas. La identidad de género no debe confundirse con el sexo biológico, ni con la expresión de género (Organización Mundial de la Salud (OMS), 2018).
- ✧ **Perspectiva de género:** es un enfoque analítico y conceptual que busca comprender las relaciones de poder y las diferencias sociales y culturales basadas en el género. También reconoce que los roles, comportamientos, expectativas y oportunidades no están determinados biológicamente, sino que son construidos socialmente (Instituto Carl Rogers, 2023).
- ✧ **Equidad de género:** principio que promueve la justicia y la igualdad de oportunidades para todas las personas, sin importar su género. La equidad de género implica eliminar las barreras que limitan el acceso y participación de mujeres y otros grupos subrepresentados en diversas áreas, incluyendo la tecnología y la inteligencia artificial (UNESCO, 2021).
- ✧ **Perspectiva interseccional:** marco conceptual y metodológico para visibilizar cómo se entrecruzan, superponen y potencian múltiples ejes de opresión o exclusión — como el género, la raza o la clase social— en la experiencia de vida de las personas. Desde esta perspectiva, se reconoce que la identidad es el resultado de múltiples dimensiones y características que configuran experiencias diferenciadas y producen formas específicas, simultáneas y acumulativas de discriminación y desigualdad (Viveros Vigoya, 2023).
- ✧ **Sesgo de género:** es la tendencia a favorecer o perjudicar a un género sobre otro de manera consciente o inconsciente, basándose en estereotipos, roles o normas sociales tradicionales. En el contexto de la inteligencia artificial, el sesgo de género se refiere a la discriminación o tratamiento desigual que se introduce en los sistemas debido a datos sesgados o algoritmos que refuerzan inequidades históricas y sociales relacionadas con el género (Bolukbasi, Chang, Zou, & Saligrama, 2016).
- ✧ **Auditoría de sesgo de género:** proceso de revisión de los sistemas de IA para identificar y corregir sesgos que afectan de manera desproporcionada a mujeres o personas de género diverso. Este proceso es clave para evitar que la IA perpetúe desigualdades preexistentes (Raji & Buolamwini, 2019)

- ☛ **STEM:** acrónimo en inglés que engloba las áreas de conocimiento de ciencia, tecnología, ingeniería y matemáticas, las cuales son clave para el desarrollo de capacidades en el ámbito tecnológico, incluida la IA.
- ☛ **Brecha digital de género:** se refiere a las diferencias en el acceso, uso y calidad de uso de tecnologías digitales entre hombres y mujeres, que afectan la capacidad de las mujeres, especialmente en países del sur global, para acceder al ecosistema digital y beneficiarse de los avances tecnológicos (Fondo Monetario Internacional, 2022).

1

¿Por qué necesitamos una guía para prevenir los sesgos de género en sistemas de inteligencia artificial?

La inteligencia artificial (IA) se presenta a menudo como una solución técnica y objetiva para gestionar interacciones digitales cotidianas, como la personalización de contenidos en redes sociales, la selección automatizada de perfiles en plataformas de empleo o la asignación de créditos mediante modelos predictivos. Su uso también se expande rápidamente en el sector público, desde el análisis de datos para la formulación de políticas públicas más efectivas y la simulación de escenarios para evaluar impactos ambientales o financieros, hasta la optimización de la inversión estatal (OECD, 2024).

En ese contexto de expansión global, Colombia ha mostrado avances significativos en la adopción de IA. Según el *Government AI Readiness Index* (Oxford Insights, 2024) —que evalúa la capacidad de los gobiernos para implementar estas tecnologías— el país ocupa la posición 55 entre 188 naciones y se ubica en el cuarto lugar a nivel latinoamericano, lo cual refleja un interés creciente del Estado por impulsar el desarrollo e implementación de modelos y sistemas de IA.

Por su parte, el Índice Latinoamericano de Inteligencia Artificial (ILIA) clasifica a Colombia como un país “adoptante”, destacando el uso extendido de estas tecnologías en distintos sectores y los avances de política pública en la materia (CENIA, 2024). El uso creciente de IA, tanto en el ámbito público como privado, no solo amplía las posibilidades de aplicación de estas tecnologías, sino que también expone tensiones de fondo sobre su diseño y uso.

Detrás de los sistemas de inteligencia artificial hay decisiones humanas: ¿qué criterios definen la visibilidad de un contenido, el perfil de un candidato “ideal” en una plataforma de empleo o las trayectorias de vida que se consideran relevantes al momento de decidir quién puede acceder a un crédito? ¿Con qué datos y cómo se ponderan los criterios para asignar recursos en salud o educación? Estas cuestiones no son puramente técnicas, nos invitan a reconocer que la tecnología no es neutral¹; está hecha por personas situadas en contextos sociales, culturales y económicos, cuyos valores se trasladan —consciente o inconscientemente— a las tecnologías que diseñan (Suchman, 2002; FRA, 2022).

Esa dimensión humana del diseño tecnológico implica que al construir sistemas de IA se establecen prioridades y se definen criterios de inclusión y exclusión. En este proceso, los sesgos y las desigualdades sociales pueden introducirse, filtrarse y amplificarse, permeando

¹ De acuerdo con la socióloga feminista Wajcman (2006), las tecnologías son productos socio-técnicos contruidos a partir de las relaciones sociales que las producen y utilizan. De esta manera, las tecnologías no son puramente técnicas o sociales, sino redes complejas que combinan artefactos, personas, organizaciones, significados culturales y conocimiento; por lo tanto, están codificadas con significados de género que inciden tanto en su diseño como en su uso.

el diseño, entrenamiento y despliegue de los sistemas. Así las cosas, la inteligencia artificial no solo refleja el mundo social, sino que también puede reforzar las exclusiones y formas de discriminación ya existentes. Un abordaje reflexivo de estos asuntos es lo que marca la diferencia entre el uso de herramientas digitales para generar oportunidades o para restringirlas.

Existe evidencia suficiente sobre cómo la IA incorpora sesgos de género (*gender bias*): prejuicios o estereotipos que se filtran en la tecnología y generan resultados que favorecen o perjudican a las personas según su sexo biológico o identidad de género (ONU Mujeres, 2025; Viteri & Gómez Pineda, 2023).

Este fenómeno es consecuencia de desigualdades estructurales que atraviesan el ámbito tecnológico, entre estas la subrepresentación de mujeres en la industria de la IA asociada a la brecha de género en áreas STEM², la falta de datos representativos y el uso de algoritmos entrenados sobre patrones de exclusión.

Entre los efectos negativos de estos sesgos se encuentra que: producen una menor calidad de los servicios para las mujeres, asignación injusta de oportunidades, refuerzo de estereotipos y prejuicios, así como tratamiento a través de los modelos de IA que puede resultar ofensivo o despectivo (UNESCO, 2024). Además, los sesgos de género no solo afectan los resultados que generan los sistemas, sino también las oportunidades de participación y toma de decisiones en el diseño de tecnologías que impactan la vida cotidiana. No se trata de fallas técnicas menores; en cambio, son expresiones de desigualdades estructurales que pueden reforzar y ampliar brechas históricas.

En Colombia, ignorar estos sesgos pone en riesgo el ejercicio pleno de derechos y profundiza la exclusión de mujeres en su diversidad —racializadas, rurales, transgénero, en situación de pobreza, con discapacidad— en sectores como la inclusión financiera, el empleo, la justicia y la participación pública³. No adoptar buenas prácticas a medida que la IA se integra en la toma de decisiones públicas y privadas, significa normalizar la discriminación algorítmica, lo que podría generar un modelo de innovación tecnológico que silencia las diferencias y amplifica la inequidad.

Es ese el contexto que motiva la construcción de esta guía de buenas prácticas para la detección, prevención y mitigación de sesgos de género en los sistemas de IA, para promover la equidad de género en el ecosistema digital. Con este propósito, el primer apartado de esta

² La representación de las mujeres en la industria de la IA es solo el 30% (UNESCO, 2024).

³ En Colombia se han implementado modelos de IA en asuntos relacionados con: transparencia y anticorrupción en la participación pública; seguridad ciudadana; protección de poblaciones en riesgo (defensores/as de derechos humanos) y prevención de la trata de personas (Mójica Muñoz, 2022). De manera semejante, en Bogotá las aplicaciones de IA reflejan una orientación pública y social, centrada en la eficiencia de los servicios, la participación ciudadana y la inclusión laboral (Alta Consejería Distrital de TIC, 2023).

guía presenta una contextualización de las iniciativas de política pública nacional y los marcos de referencia global que promueven una IA equitativa; posteriormente se abordan los objetivos de la guía y los públicos a los que está dirigida.

Después, se desarrolla un encuadre conceptual más detallado sobre los sesgos de género en los sistemas de IA y, finalmente, se presenta un conjunto de buenas prácticas, un catálogo de herramientas y un árbol de decisión diseñados para apoyar a las personas que participan en el ecosistema de IA en la incorporación de enfoques más inclusivos y sensibles al género en sus procesos y desarrollos tecnológicos.

Deseamos que esta guía y sus herramientas sean útiles en la misión colectiva de construir un sistema de IA equitativo, porque **lo que está en juego no es únicamente la calidad de nuestras tecnologías, sino también el tipo de sociedad que queremos construir con ellas.**

El compromiso de Colombia por impulsar una transformación digital con equidad de género

Para hacer frente a los riesgos y desafíos relacionados con los sesgos de género en los sistemas de IA, esta guía de buenas prácticas se articula con marcos normativos y lineamientos de política pública orientados a mitigar los sesgos de género en los sistemas de inteligencia artificial y promover un desarrollo tecnológico más justo e inclusivo. Se trata de una acción coherente con el compromiso del país de avanzar hacia una transformación digital que contribuya a cerrar brechas y promover la equidad de género.

En este proceso, uno de los pilares lo constituye el **CONPES 4080: Política Pública de Equidad de Género para las Mujeres** (DNP, 2022) que establece una hoja de ruta para promover la igualdad de oportunidades y garantizar los derechos de las mujeres en diversos ámbitos. Específicamente, esta guía de buenas prácticas materializa la Acción 1.17 de dicho CONPES, cuyo objetivo es diseñar e implementar estrategias para la identificación y prevención de sesgos de género en el desarrollo, entrenamiento, validación y aplicación de sistemas de inteligencia artificial y otras tecnologías digitales.

Esta política se alinea con el **Plan Nacional de Desarrollo (PND) 2022-2026 “Colombia Potencia Mundial de la Vida”** (DNP, 2023), que subraya la persistencia de desigualdades y discriminaciones basadas en roles y estereotipos de género, y la necesidad de reconocer a las mujeres como agentes de los cambios sociales, económicos y culturales que promueve el Plan. Así mismo, el PND hace énfasis en las transformaciones necesarias para superar las violencias y desigualdades basadas en género, orientación sexual e identidades de género diversas, lo cual resuena con la necesidad de una IA libre de sesgos, en tanto estas tecnologías pueden replicar o intensificar las violencias que buscan erradicarse.

Por su parte, la **Estrategia Nacional Digital (END)** (DNP, 2024) refuerza el llamado a eliminar las brechas tecnológicas mediante políticas digitales inclusivas que promuevan la equidad como principio transversal. En particular, esta Estrategia prioriza el desarrollo y uso de tecnologías emergentes, como la inteligencia artificial, con un enfoque de género, reconociendo que su implementación equitativa es clave para avanzar en la inclusión digital y social.

En diálogo con los marcos éticos promovidos a nivel global, el Gobierno nacional, a través del **Marco Ético para la Inteligencia Artificial en Colombia** (Gobierno Nacional de Colombia, 2021) define principios fundamentales para el diseño, desarrollo y uso responsable de estas tecnologías en el país. Este marco promueve un enfoque centrado en las personas, con énfasis en la transparencia, la seguridad, la equidad y el respeto por los derechos humanos. Además, busca orientar a los diferentes actores (tanto del sector público como privado) en la toma de decisiones informadas, anticipando riesgos y promoviendo un desarrollo tecnológico que responda a las necesidades del país.

En el ámbito judicial, este enfoque ha sido respaldado por la **Sentencia T-323 de 2024** (Corte Constitucional de Colombia, 2024), por la cual se dictan los criterios orientadores para el adecuado uso de la inteligencia artificial (IA). Esta sentencia refuerza el Marco Ético para la IA en Colombia al hacer énfasis en el respeto de los derechos fundamentales y en la explicabilidad en el uso de IA, con el fin de evitar que las decisiones automáticas discriminen por género.

En materia de instrumentos de política pública directamente enfocados en IA, la **Hoja de Ruta en Inteligencia Artificial** (Ministerio de Ciencia, Tecnología e Innovación, 2024) establece lineamientos específicos para el desarrollo de una IA transparente y ética, orientada a minimizar sesgos discriminatorios en su diseño y aplicación. Sobre esa base, el **CONPES 4144 - Política Nacional de Inteligencia Artificial** (DNP, 2025) propone generar capacidades para el desarrollo y aprovechamiento ético y sostenible de la IA, impulsando la transformación social y económica del país.

En particular, en materia de prevención de sesgos, el CONPES 4144 establece que el Departamento Administrativo Nacional de Estadística (DANE) liderará el diseño e implementación de una estrategia para fortalecer la inclusión del enfoque diferencial e interseccional en el ciclo de vida de los datos, esto en pro de garantizar la representatividad de diversos grupos poblacionales en el marco de entrenamiento de modelos de IA.

En conjunto, estas políticas e iniciativas trazan un camino para Colombia en la construcción de un futuro donde la tecnología, con la IA en la vida cotidiana, impulse la equidad. Esta guía se inserta como una herramienta clave para alcanzar ese objetivo, brindando herramientas prácticas que contribuyan a que la transformación digital sea justa y beneficie a todas las personas.

Diálogos globales: marcos internacionales para una IA ética e inclusiva

El compromiso de Colombia con una inteligencia artificial más equitativa se enmarca en una conversación global cada vez más relevante sobre la necesidad de promover tecnologías responsables.

En particular, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) y la Organización para la Cooperación y el Desarrollo Económicos (OCDE) han hecho un trabajo importante al promover principios para una IA centrada en el ser humano, enfatizando la necesidad de prevenir la discriminación y promover la equidad de género en el diseño, desarrollo y despliegue de estas tecnologías. Estos marcos son fundamentales porque ofrecen un lenguaje común y estándares mínimos que inspiran y orientan acciones como la creación de esta guía.

Por un lado, están las **Recomendaciones sobre la Ética de la Inteligencia Artificial de la UNESCO** (2021), adoptadas por 193 países, incluido Colombia (marzo de 2022). Este es el primer instrumento normativo global que establece principios y valores para una IA responsable, y además, define áreas de acción concretas. Entre sus objetivos está el de “proteger, promover y respetar los derechos humanos y las libertades fundamentales, la dignidad humana y la igualdad, incluida la igualdad de género”.

La UNESCO establece que los gobiernos deberían “*velar porque se aproveche al máximo el potencial de las tecnologías digitales y la inteligencia artificial para contribuir al logro de la igualdad de género, y deben garantizar que los derechos humanos y las libertades fundamentales de las niñas y las mujeres, así como su seguridad e integridad, no se vulneren en ninguna etapa del ciclo de vida del sistema de IA*” (Recomendación #87).

Por su parte, la OCDE, a cuyas **Recomendaciones sobre Inteligencia Artificial** (2019) a la que se adhirió Colombia en mayo de 2019, plantea que “*las partes interesadas deben participar de manera proactiva en una gestión responsable de la IA confiable con el fin de (...) promover la inclusión de poblaciones vulnerables*”. Llevar estos principios a la práctica en cada país implica un trabajo adicional para adaptarlos y crear herramientas útiles y aplicables.

Referentes semejantes a esta guía de buenas prácticas

A nivel internacional, el Instituto de Liderazgo Digital (NDLI) - *NetHope* cuenta con recursos como el **Gender Equitable AI Toolkit** (2021), que ofrece lineamientos valiosos para incorporar la perspectiva de género en el diseño y aplicación de soluciones de IA, especialmente en contextos humanitarios y de desarrollo. Este recurso integra metodologías participativas y mecanismos de auditoría algorítmica para asegurar que los modelos desarrollados sean justos y representativos; además ofrece listas de verificación y preguntas orientadoras que facilitan la implementación práctica del enfoque de género en IA.



Desde el *Center for Financial Inclusion* y USAID se desarrolló “**Una guía de gestión de riesgos para inversores de impacto**” (Álvarez & Rizzi, 2023), enfocada en proporcionar herramientas prácticas a inversionistas y empresas que utilizan inteligencia artificial. Su objetivo es promover el uso equitativo de la IA en las finanzas inclusivas, ayudando a identificar y mitigar los sesgos que pueden afectar negativamente a grupos vulnerables, especialmente mujeres, y garantizar que las decisiones algorítmicas sean responsables.

Por su parte, el ***Red Teaming Artificial Intelligence for Social Good*** de la UNESCO (2025) es una guía práctica diseñada para democratizar el acceso a herramientas de evaluación crítica de los modelos de inteligencia artificial generativa (Gen AI). A través de ejercicios estructurados, este recurso permite que organizaciones, gobiernos, comunidades y ciudadanía en general puedan identificar sesgos, estereotipos y daños potenciales — especialmente aquellos que afectan de manera desproporcionada a mujeres y niñas, como la violencia de género facilitada por tecnologías (TFGBV).

La metodología de *Red Teaming* propuesta permite simular situaciones reales donde los sistemas de IA pueden generar resultados dañinos, ya sea por sesgos incrustados en los datos o por su uso malintencionado. El Playbook incluye instrucciones paso a paso, ejemplos de desafíos temáticos, formatos de participación (experta o pública), recursos para el análisis de resultados y un enfoque centrado en la seguridad psicológica de los participantes. Esta herramienta no solo contribuye a fortalecer la rendición de cuentas y la gobernanza ética de la IA, sino que también promueve una participación activa de los grupos más afectados en la detección de riesgos, generando evidencia para impulsar cambios normativos, técnicos y culturales hacia una IA más justa y equitativa.

Finalmente, en el ámbito nacional, vale la pena mencionar el documento **IA Inclusiva - herramientas para el uso responsable de Inteligencia Artificial** (C4IR.CO), elaborado por el Centro para la Cuarta Revolución Industrial de Colombia (C4RI). Se trata de una guía dirigida a tomadores de decisión y equipos técnicos interesados en entender las limitaciones de la IA y las alternativas para superarlas a través de ocho herramientas para prevenir y mitigar sesgos en sistemas de IA.

2

¿Para qué sirve esta guía de buenas prácticas y cómo puede apoyar a quienes integran el ecosistema de IA?

El propósito principal de esta guía es impulsar en el ecosistema de inteligencia artificial nacional un cambio en la forma en la que se conciben, diseñan, implementan y evalúan los modelos y sistemas de IA. Para ello, ofrece un conjunto de buenas prácticas y herramientas concretas que permiten incorporar la equidad de género a través de la detección y mitigación de sesgos de género en el ciclo de vida de estas tecnologías. Además, la apropiación de la guía y de las herramientas que se incluyen tiene el potencial de catalizar otras transformaciones significativas como:

- ✔ **Toma de decisiones públicas y privadas más justas:** al integrar la evaluación de sesgos de género como un estándar, se promueven decisiones más equitativas en la contratación de tecnologías, el diseño de servicios públicos digitales y las estrategias de innovación empresarial.
- ✔ **Desarrollo tecnológico con responsabilidad social:** fomenta una cultura de diseño y desarrollo de IA que internaliza la equidad desde la concepción, reduciendo la probabilidad de impactos discriminatorios y mejorando la calidad y pertinencia de las soluciones tecnológicas para toda la población.
- ✔ **Fortalecimiento del ecosistema de innovación:** la conformación de equipos diversos en términos de género y el uso de datos más representativos potencian la capacidad de innovación, al ampliar los enfoques, experiencias y perspectivas involucradas en el desarrollo de soluciones tecnológicas basadas en IA. Esto permite generar respuestas más pertinentes, inclusivas y adaptadas a la diversidad de realidades del país.
- ✔ **Avance en la agenda de equidad de género:** contribuye directamente a los objetivos nacionales de cierre de brechas de género, no solo en el sector tecnológico, sino en todos los ámbitos donde la IA tenga incidencia.
- ✔ **Posicionamiento de Colombia como referente de equidad:** la implementación efectiva de esta guía puede posicionar al país como un líder regional en el desarrollo y uso ético y equitativo de la IA.

¿A quiénes convoca esta guía?

La transformación hacia una IA con equidad de género es un esfuerzo colectivo, de allí que esta guía, así como el Catálogo de Herramientas y Árbol de Decisión, estén dirigidas a todas las personas que desempeñan roles esenciales en el ecosistema de IA:

-  **Equipos de desarrollo de IA (ingenieras/os, científicas/os de datos, diseñadoras/es UX/UI)⁴:** esta guía ofrece recursos para integrar la perspectiva de género desde la concepción de un sistema, seleccionar datos representativos, auditar algoritmos y diseñar interfaces inclusivas.
-  **Entidades públicas y organismos reguladores:** la guía proporciona criterios para la contratación de soluciones basadas en IA con enfoque de género, insumos para el diseño de políticas públicas y mecanismos para la evaluación del impacto de la IA en términos de equidad de género.
-  **Sector privado (empresas tecnológicas, startups, consultoras):** la guía es útil para desarrollar productos y servicios de IA que no solo sean competitivos, sino también éticos y equitativos, abriendo oportunidades en mercados más conscientes.
-  **Academia e investigadores/as:** sirve como base para desarrollar nuevas líneas de investigación, para la formación de profesionales con sensibilidad a la inclusión de enfoque de género en IA y para el desarrollo de soluciones innovadoras que aborden los sesgos.
-  **Organizaciones de la sociedad civil, colectivos de defensa de los derechos de las mujeres y la equidad de género:** ofrece herramientas para auditar sistemas, incidir en políticas y asegurar que los derechos de las mujeres y personas de género diverso sean protegidos en el entorno digital.
-  **Personas usuarias finales:** aunque no siempre se perciba, todas y todos interactuamos con sistemas de IA. Esta guía busca empoderar a las personas para entender sus derechos, identificar posibles sesgos y participar activamente en la mejora de las tecnologías que usamos.
-  **Medios de comunicación y periodistas especializados en tecnología:** la guía ofrece insumos para informar con responsabilidad sobre los impactos sociales de la IA y el imperativo de la equidad de género.

¿Qué contiene esta guía?

Para facilitar esta transformación y asegurar que quienes integran el ecosistema digital nacional cuenten con el apoyo necesario en cada etapa del ciclo de vida de la IA, esta guía está diseñada como un recurso práctico que incluye:

Buenas Prácticas: acciones y recomendaciones específicas para integrar la perspectiva de género a lo largo de todo el ciclo de vida de los sistemas de IA. Desde la planificación y diseño

⁴ User Experience (Experiencia de Usuario) y User Interface (Interfaz de Usuario), respectivamente.

inicial, pasando por la recopilación y tratamiento de datos, la creación de modelos, las pruebas y validación, el despliegue, la supervisión continua, hasta el desmantelamiento.

De manera complementaria, la guía se potencia con dos herramientas diseñadas para apoyar la detección, prevención y mitigación de sesgos de género:

Catálogo de Herramientas: reúne recursos prácticos como software, librerías, frameworks, metodologías, listas de verificación y casos de uso, diseñados para identificar, prevenir y mitigar sesgos de género en la IA. Estas herramientas son de acceso libre, factibles de implementar en contextos locales, y están orientadas a promover la equidad de género y la creación de sistemas de IA más justos. Además, el catálogo de herramientas proporciona guías técnicas y ejemplos aplicados que facilitan el diseño y la evaluación responsable de soluciones tecnológicas.

Árbol de Decisión: es una herramienta interactiva estructurada y diseñada para guiar a las personas usuarias en la identificación y abordaje de posibles sesgos de género en cada etapa del desarrollo, adquisición o implementación de sistemas de inteligencia artificial (IA). Facilita la toma de decisiones informadas al ofrecer criterios claros sobre datos, metodologías y herramientas, además de proponer acciones específicas para mitigar los riesgos detectados. Su enfoque interactivo garantiza la incorporación de principios éticos y de equidad en todas las fases del ciclo de vida de la IA, promoviendo soluciones responsables e inclusivas.

Figura 1 Guía para una IA equitativa en Colombia



Fuente: elaboración propia con apoyo de Napkin

¿Cómo se reproducen desigualdades de género en sistemas de inteligencia artificial?

Los sistemas de IA no son una “caja negra” que toma decisiones. Son el resultado de un proceso compuesto por varias fases, y cada una es un punto crítico donde los sesgos de género pueden surgir, pero también corregirse.

- 1. Planificación:** en esta etapa se definen los objetivos del sistema, a partir de los usos y contexto de despliegue. Las decisiones tomadas en esta etapa definen el propósito del sistema, a quién servirá y cómo lo hará. Si no se incluye desde el inicio un enfoque diferencial, se corre el riesgo de excluir necesidades específicas de grupos poblacionales. Si no se formulan preguntas con un enfoque de equidad desde el inicio, los sistemas pueden invisibilizar a mujeres y personas con identidades de género diversas, así como omitir problemáticas percibidas por estos grupos. También es común que los equipos de desarrollo no incluyan diversidad en su composición, lo que puede limitar la identificación temprana de sesgos y la exclusión de realidades y contextos.
- 2. Diseño y desarrollo:** se seleccionan los datos y se entrenan los modelos. Es decir, en esta fase se toman decisiones sobre las fuentes de datos y el diseño algorítmico, ambos profundamente sensibles al sesgo. Muchos sistemas de IA se entrenan con datos históricos que reflejan desigualdades de género existentes, por lo que pueden “aprender” a reproducirlas. Además, la falta de disponibilidad de datos, así como la no desagregación por sexo biológico e identidad de género o de otras variables que permiten entender la interseccionalidad (raza, ubicación geográfica, ingresos) impide detectar patrones discriminatorios. También es común que los modelos favorezcan correlaciones que refuerzan estereotipos (por ejemplo, asociar ciertos roles ocupacionales a un género específico).
- 3. Validación y prueba:** se evalúa cómo funciona el sistema. Es crucial incluir métricas de equidad y probar el sistema con poblaciones diversas para detectar diferencias en los resultados. Un sistema puede funcionar bien en términos generales, pero tener impactos desproporcionadamente negativos sobre ciertos grupos. Si las pruebas no incluyen escenarios diferenciados o conjuntos de datos que representen adecuadamente a mujeres y poblaciones diversas, se corre el riesgo de validar modelos que perpetúan sesgos. La ausencia de métricas específicas de equidad de género impide identificar errores sistemáticos como tasas de falsos positivos o negativos diferenciadas por género.
- 4. Despliegue o implementación:** es el momento en que el sistema empieza a interactuar con personas reales. Si no hay protocolos para retroalimentación o corrección, los sesgos pueden pasar desapercibidos y replicarse a gran escala.

Durante esta etapa, los efectos de un sistema sesgado dejan de ser teóricos y comienzan a impactar decisiones reales —como el acceso a beneficios, servicios, empleos o justicia—. Sin mecanismos robustos de monitoreo, los sesgos pueden no solo persistir, sino amplificarse. Además, muchas veces no se informa a las personas usuarias cómo pueden reportar errores, ni se establecen vías accesibles para impugnar decisiones automatizadas, lo que agrava los riesgos para poblaciones históricamente discriminadas.

5. **Monitoreo y supervisión:** se observa el comportamiento del sistema en el tiempo y se hacen ajustes si se detectan sesgos emergentes, teniendo en cuenta que los contextos cambian y los sistemas de IA pueden adaptarse de formas imprevistas. Un modelo que parecía equitativo al inicio puede derivar en prácticas discriminatorias si no se supervisa su desempeño con perspectiva de género. Además, los sesgos pueden no manifestarse de inmediato, sino emerger gradualmente. Por tanto, es importante implementar auditorías algorítmicas periódicas, con indicadores de impacto desagregados por género, así como mecanismos de acción correctiva.
6. **Desmantelamiento o reemplazo:** cuando un sistema es obsoleto o perjudicial, debe retirarse de forma responsable, evaluando el daño potencial causado por sesgos no corregidos y documentar las lecciones aprendidas. El cierre de ciclo suele ser subestimado en términos de equidad. Sin una revisión crítica del impacto diferenciado que tuvo el sistema, se pierden oportunidades valiosas para mejorar futuros desarrollos. Además, los sistemas de IA pueden dejar efectos residuales —por ejemplo, bases de datos que conservan sesgos o decisiones automatizadas sin posibilidad de reversión—. Documentar los aprendizajes en clave de género y generar mecanismos de reparación o compensación, cuando sea necesario, es fundamental para una gobernanza ética de la IA.

¿Cuáles son los sesgos de género en sistemas de IA?

Los sistemas de IA pueden verse afectados por diversos tipos de sesgo, como los raciales, culturales, socioeconómicos y etarios o generacionales. Esta guía enfatiza en los sesgos de género, por lo que este apartado profundiza en su tipificación y características. Asimismo, la evidencia se ha enfocado en los sesgos que producen resultados negativos de forma directa⁵, pues pueden tener impactos desproporcionados sobre mujeres y personas con identidades de género diversas (UNESCO, IRCAL, 2024; UNESCO, 2024).

⁵ No todos los sesgos de género en sistemas de inteligencia artificial se manifiestan como desventajas explícitas; algunos pueden parecer “favorables”, pero igualmente reproducen desigualdades. Por ejemplo, en el sector financiero, un sistema automatizado puede asignar límites de crédito más bajos a las mujeres con base en datos históricos que indican menor riesgo financiero. Aunque esta medida aparenta ser prudente, termina restringiendo su autonomía económica, reforzando estereotipos y

En términos generales, el Marco de Gestión de Riesgos de IA del *National Institute of Standards and Technology*, distingue tres categorías de sesgo en IA: **sesgos sistémicos**, que pueden estar presentes en los conjuntos de datos de IA, en las normas, prácticas y procesos a lo largo del ciclo de vida de la IA, así como en las personas que usan estos sistemas; **sesgos computacionales y estadísticos**, los cuales pueden encontrarse en los datos y en los procesos algorítmicos, y en general provienen de errores sistemáticos debido a muestras no representativas; y **sesgos cognitivos humanos**, que hacen referencia a cómo un individuo o grupo percibe la información de un sistema de IA para tomar decisiones, o cómo piensan sobre los propósitos y funciones de dicho sistema (National Institute of Standards and Technology, 2023).

De manera más específica, otras fuentes han tipificado los sesgos en sistemas de IA como se muestra en la tabla a continuación:

Tabla 1 Principales sesgos en sistemas de inteligencia artificial

<i>Tipo de sesgo</i>	<i>Descripción</i>	<i>Ejemplo</i>
Sesgo de datos	Se presenta cuando los algoritmos se entrenan con datos no representativos de la diversidad humana, lo cual genera decisiones sesgadas porque tienden a basarse en la experiencia de pocos grupos. En el ámbito de la salud, esto afecta diagnósticos y tratamientos al no considerar diferencias fisiológicas entre géneros (Buolamwini & Gebru, 2018); en el ámbito laboral, refuerza la preferencia por perfiles masculinos, perpetuando la exclusión de mujeres y minorías en ciertos roles históricos (O'Neil, 2016).	Estudios sobre un sistema de reconocimiento facial en seguridad pública en Reino Unido, revelaron que el sistema tenía una mayor tasa de error al identificar a mujeres y personas negras. La falta de diversidad en los datos de entrenamiento contribuyó con esos errores., generando problemas de discriminación y vigilancia desproporcionada. Esto a su vez, planteó preocupaciones sobre la posible criminalización injusta de mujeres y minorías étnicas (Grother, Ngan, & Hanaoka, 2019) (Murgia, 2019).
Sesgo de confirmación	Este sesgo ocurre cuando los algoritmos priorizan información que refuerza estereotipos preexistentes, como asociar roles	En 2018, Amazon desarrolló un sistema de reclutamiento basado en IA para optimizar la selección de personal. El algoritmo penalizaba los currículums que contenían palabras o

negándoles acceso equitativo a recursos. Algo similar puede ocurrir en el ámbito de la salud: si un sistema prioriza automáticamente las citas médicas preventivas para mujeres, bajo la lógica de que ellas suelen acudir más al sistema de salud, podría terminar desplazando a hombres u otros grupos que, precisamente por tener menor acceso, podrían necesitar una mayor focalización. En ambos casos, lo que parece una ventaja o un “sesgo positivo” encubre decisiones basadas en género que distorsionan la equidad, perpetúan sesgos y reducen la confianza en los sistemas automatizados.

	laborales tradicionalmente masculinos, lo cual perpetúa estas ideas preconcebidas en procesos de contratación y selección de personal (Buolamwini & Gebru, 2018).	actividades estereotípicamente asociadas a actividades o roles femeninos, ya que fue entrenado con datos históricos de currículums en los que predominaban candidatos masculinos. Este sesgo redujo las oportunidades laborales para mujeres en el sector tecnológico (Dastin, 2018).
Sesgo de interacción	La interacción de los sistemas de IA varía según el género, reflejando y reforzando estereotipos tradicionales en las respuestas de asistentes virtuales. Por ejemplo, los sistemas de IA como Alexa y Siri utilizan voces femeninas en roles serviles y voces masculinas en contextos autoritarios, lo cual contribuye a la idea tradicional de género asociada a la autoridad y el servicio (UNESCO, 2024).	Los principales asistentes virtuales como Siri, Alexa y Google Assistant se diseñaron inicialmente con voces y personalidades femeninas, al mismo tiempo se programaron para que respondieran de manera sumisa, con lo cual se refuerza la percepción de las mujeres ocupan roles de servidumbre y asistencia (Kershner, 2018) (West, Kraut, & Chew, 2019).
Sesgo de uso	Las barreras estructurales dificultan el acceso equitativo de mujeres y personas LGBTQ+ a los sistemas de IA. La falta de acceso a dispositivos tecnológicos y la exclusión en los datos de entrenamiento limitan las oportunidades de estos grupos para beneficiarse plenamente de los avances tecnológicos, lo que amplía las brechas de género existentes (UNESCO, 2024).	
Sesgo de retroalimentación	Se produce cuando los resultados de un sistema de IA influyen en los datos futuros que procesará el mismo sistema, creando un ciclo que refuerza los sesgos preexistentes. Este ciclo amplifica las desigualdades en áreas como la contratación, donde se perpetúa la exclusión de las mujeres, o en salud, al no tener en cuenta sus necesidades específicas en los	En India se implementaron sistemas de IA para evaluar la solvencia crediticia de solicitantes de préstamos. Los algoritmos clasificaban a las mujeres como más riesgosas debido a la base de datos histórica que reflejaba una baja participación de mujeres en la obtención de créditos,

**Sesgo contextual
o de aplicación**

tratamientos y diagnósticos (Raji & Buolamwini, 2019).

Aparece cuando los sistemas de IA se aplican en contextos diferentes a los previstos, generando discriminación de género al no ajustarse a las necesidades socioculturales específicas de mujeres y personas LGBTQ+. Los algoritmos que no consideran diversas realidades socioculturales terminan exacerbando desigualdades y limitando la efectividad de los sistemas en diferentes entornos (O'Neil, 2016).

Fuente: DNP-Betagroup, 2024.

4 ¿Cómo podemos promover la equidad de género en los sistemas de inteligencia artificial?

Este apartado presenta un conjunto de buenas prácticas que se basan en principios éticos y recomendaciones de organismos internacionales como la UNESCO (2021) y la OCDE (2024), así mismo se incluyen ejemplos concretos de su aplicación en diversos contextos según las fases del ciclo de vida de los sistemas de IA. El objetivo es mostrar cómo la implementación de estas prácticas puede fortalecer la diversidad, la inclusión y la equidad en el diseño de tecnologías de IA.

Buenas prácticas para el desarrollo de sistemas de IA

La implementación de buenas prácticas en la **planificación y diseño de la IA** es esencial para mitigar sesgos de género desde el inicio. Una planificación sólida y directrices claras permiten detectar y prevenir desigualdades estructurales evitando su reproducción en los modelos.

Tabla 2 Buenas prácticas en la etapa de planificación y diseño

Buena práctica	Ejemplo de aplicación
<i>Implementar directrices inclusivas en el diseño para asegurar la representación justa y precisa de las comunidades de género diverso, incluidas las personas LGBTQ+, en todas las fases del desarrollo de productos y servicios impulsados por IA.</i>	GLAAD, en colaboración con Google, ha creado "All In", un conjunto de recursos de marketing inclusivo que ofrece directrices detalladas para representar de manera auténtica a comunidades subrepresentadas, incluidas las personas LGBTQ+. Estas directrices buscan ayudar a los profesionales del marketing a evitar estereotipos y a reflejar con precisión la diversidad de la comunidad LGBTQ+ (GLAAD, 2021).
Enfoque: prevención de sesgos de género	
<i>Cuestionar la neutralidad de los datos y el poder detrás de su recolección desde el inicio del proyecto.</i>	Estudio "Gender Shades", en 2018, Joy Buolamwini y Timnit Gebre evaluaron tres sistemas comerciales de reconocimiento facial (de IBM, Microsoft y Face++) y descubrieron que las tasas de error en la clasificación de género llegaban hasta un 34,7 % para mujeres de piel más oscura, frente a menos del 1 % para hombres de piel clara, debido a que

Enfoque: prevención de sesgos de género	los conjuntos de datos de entrenamiento estaban compuestos en más de un 80 % por rostros claros y masculinos. Tras visibilizar esta asimetría, publicaron un nuevo banco de imágenes (Pilot Parliaments Benchmark) equilibrado por género y tono de piel y establecieron recomendaciones para recolección y evaluación de datos inclusivos. (Joy Buolamwini and Timnit Gebru, s.f., pág. 15).
<i>Avanzar hacia la estandarización e incorporación de variables o atributos de género en las bases de datos nacionales (registros administrativos).</i>	Censo de Canadá 2021, por primera vez el censo canadiense incluyó dos preguntas diferenciadas: “¿Cuál fue el sexo asignado a la persona al nacer?” y “¿Cuál es el género de la persona?” Con opciones de “masculino”, “femenino” y “especifique” para cubrir identidades no binarias. Además, se publicó una Guía de Referencia para “Edad, sexo al nacer y género” que estandariza definiciones, clasificaciones y metadatos, permitiendo mantener comparabilidad histórica y, al mismo tiempo, garantizar la visibilidad de poblaciones trans y no binarias (Canada, 2022).
Enfoque: prevención de sesgos de género	Gender Budgeting: Step-by-step toolkit de EIGE - Guidance for mainstreaming gender into the EU Funds. Obligación de realizar EIDG para todos los programas financiados, identificando riesgos de exclusión y proponiendo variables de mitigación (p. ej. cobertura de servicios en zonas rurales femeninas) antes de recolectar datos (European Institute for Gender).
<i>Realizar una Evaluación de Impacto Diferenciado de Género (EIDG) antes de definir fuentes y variables.</i>	WHO Global Health Observatory, cada serie estadística incluye un “Indicator Metadata” que señala claramente: “Sex: male, female, both sexes”, con notas sobre población cubierta y ajustes estadísticos. El diccionario PDF detalla cómo se imputan datos faltantes por sexo y qué fuentes (censos, encuestas) se usaron (World Health).
Enfoque: prevención de sesgos de género	
<i>Añadir a cada conjunto de datos un documento de “diccionario de datos y metadatos” donde, además de variables y tipos, se especifique la dimensión de género: si la variable fue recolectada, imputada o no aplica.</i>	<i>Accenture (Sesgo de selección en IA para reclutamiento): para su propia herramienta interna de pre-selección de candidatos basada en IA, Accenture definió un indicador clave de desempeño (KPI) que exige que la tasa de oferta de empleo sea al menos igualitaria entre mujeres y hombres. Cada mes, su equipo de Analítica de Talento ejecuta un pipeline de evaluación de equidad que calcula la Disparate Impact Ratio y la Equal Opportunity Difference en los resultados del modelo. Si alguno de estos indicadores sale de rango (por debajo de 0,8 o por encima de 1,25), se detiene el despliegue y se ajusta el dataset o el algoritmo hasta corregir la desviación. Esto permite que las recomendaciones de la IA cumplan criterios de no discriminación</i>
Enfoque: prevención de sesgos de género	
<i>Establecer metas de equidad de género claras y medibles en el diseño de IA, asegurando que los resultados de los sistemas de IA respeten los principios de no discriminación y diversidad.</i>	
Enfoque: identificación y prevención de sesgos de género.	

Definir previamente las prácticas de diseño que permitan la transparencia en las decisiones algorítmicas y faciliten la comprensión de cómo los datos y modelos afectan las recomendaciones y resultados del sistema.

Enfoque: identificación de sesgos de género.

Incorporar una perspectiva de prevención de violencia de género digital desde las primeras etapas del diseño de sistemas de IA, asegurando que los productos no faciliten ni amplifiquen dinámicas de poder desiguales.

Enfoque: identificación y mitigación de riesgos de violencia de género

y que la diversidad quede reflejada en las decisiones automatizadas. (Accenture, s.f.).

Google desarrolló el What-If Tool dentro de su suite de IA para mejorar la transparencia y explicabilidad en los modelos de aprendizaje automático. Esta herramienta permite a los desarrolladores analizar cómo variables de entrada, como género o edad, afectan las salidas del modelo, identificando si existen disparidades o sesgos en los resultados (Google, s.f.).

Mediante la visualización de cómo las variables impactan las predicciones, los equipos pueden realizar ajustes en los modelos para tomar decisiones más justas. Esta práctica permite identificar y mitigar posibles sesgos de género desde la fase de diseño, promoviendo sistemas de IA más equitativos y confiables.

El documento de la UNFPA sobre la violencia de género facilitada por la tecnología (TF GBV) presenta directrices claras que abordan cómo la tecnología puede influir en la seguridad y bienestar de las mujeres y personas de identidades diversas, además de destacar las acciones necesarias para mitigar este tipo de violencia. Los principios clave incluyen la necesidad de diseño seguro y privacidad por defecto, que están estrechamente relacionados con la protección de los derechos de las personas y la seguridad digital (UNFPA, s.f.).

Fuente: DNP-Betagroup, 2024.

En la **recopilación y tratamiento de datos** en IA, la calidad y representatividad de los datos son cruciales para evitar la amplificación de sesgos y promover la equidad. Es clave garantizar datos diversos, representativos y validados, proteger la privacidad, e implementar auditorías y técnicas innovadoras que identifiquen y mitiguen sesgos desde el inicio.

Tabla 3 Buenas prácticas en la etapa de recopilación y tratamiento de datos

Buena práctica	Ejemplo de aplicación
<i>Asegurar que los conjuntos de datos utilizados reflejen una diversidad de género adecuada y que no contengan sesgos que puedan reproducir estereotipos. Esto puede lograrse</i>	Empresas tecnológicas globales como Google, Microsoft e IBM han implementado políticas para reducir los sesgos de género en sus sistemas de IA mediante auditorías algorítmicas y directrices internas que buscan evitar desigualdades. Google, por ejemplo, realiza auditorías de equidad algorítmica, analizando datos

mediante auditorías de equidad algorítmica de los datos para evaluar su representatividad y calidad.

Enfoque: prevención de sesgos de género

Documentar los orígenes y las características de los conjuntos de datos, incluyendo cómo se abordan las cuestiones de género en la recopilación y procesamiento de datos.

Enfoque: prevención de sesgos de género.

Fomentar el uso de conjuntos de datos abiertos y estándares que garanticen diversidad de género, y establecer controles para minimizar la recolección de datos sesgados.

Enfoque: prevención de sesgos de género.

Proteger la privacidad y anonimizar los datos de forma que se minimicen riesgos para las personas.

Enfoque: Prevención de sesgos de género

desagregados por género, raza y otros factores para identificar y corregir posibles tendencias discriminatorias (Google AI, 2020).

Un caso de uso denominado “CelebA Dataset” publicado por la Universidad China de Hong Kong (CUHK) muestra un ejemplo de prevención de sesgos a través de la documentación detallada de datasets. Este conjunto de datos de celebridades incluye más de 200.000 imágenes con anotaciones detalladas, donde se documenta explícitamente la distribución de género, los métodos de recopilación y etiquetado, y los atributos relacionados con el género de las personas.

La documentación incluye metadatos específicos sobre la composición demográfica, el proceso de recolección de imágenes, y cómo se abordaron las cuestiones de género en el etiquetado, haciendo que este dataset sea ampliamente utilizado en investigaciones sobre equidad y sesgos en sistemas de reconocimiento facial (Ziwei, Ping, Xiaogang, & Xiaoou, 2015).

IBM desarrolló el conjunto de datos Diversity in Faces para fomentar la diversidad y reducir el sesgo en sistemas de reconocimiento facial. Este conjunto de datos, diseñado con una amplia representación de género, etnicidad y otros atributos demográficos, permite a los desarrolladores entrenar modelos de IA más inclusivos y representativos. Al promover estándares de diversidad y establecer controles para evitar la recolección de datos sesgados, IBM busca asegurar que los modelos de IA reconozcan correctamente a personas de diversos grupos (Exponiendo.ai, 2019).

Apple emplea privacidad diferencial para proteger la información personal en aplicaciones como Siri y su sistema de autocorrección. Esta técnica agrega “ruido” a los datos recolectados, asegurando que la información individual se mantenga anónima incluso cuando se analizan patrones de uso general. Además, Apple limita el procesamiento de datos al dispositivo del usuario, lo cual significa que la mayoría de los datos no se envían a sus servidores. De este modo, Apple puede mejorar sus servicios y algoritmos de manera segura, minimizando los riesgos para la privacidad de las personas y asegurando que su información se mantenga protegida (Apple, s.f.).

Fuente: DNP-Betagroup, 2024.

La **construcción y adaptación de modelos** de IA son etapas críticas para prevenir la integración de sesgos de género en los algoritmos. Aplicar prácticas como técnicas de entrenamiento equitativas, pruebas exhaustivas y una cultura de responsabilidad en los equipos de desarrollo contribuye a garantizar modelos justos y respetuosos con la diversidad.

Tabla 4 Buenas prácticas en la etapa de construcción y adaptación de modelos

Buena práctica	Ejemplo de aplicación
<i>Aplicar técnicas de entrenamiento y ajuste de modelos que consideren la equidad de género, utilizando algoritmos que limiten el sesgo algorítmico y se adapten para minimizar los efectos de género en las predicciones.</i> Enfoque: prevención de sesgos de género.	Google aplicó técnicas de entrenamiento y ajuste en Google Translate para reducir el sesgo de género en sus traducciones automáticas. Antes, en idiomas sin pronombres de género específicos, el sistema tendía a traducir con estereotipos, como "doctor" en masculino y "enfermera" en femenino. Para solucionar esto, Google ajustó el modelo de IA para que proporcione traducciones en ambos géneros cuando el contexto es ambiguo. Este enfoque ayuda a limitar el sesgo algorítmico y asegura que las traducciones reflejen una perspectiva más equitativa, ofreciendo opciones de traducción tanto en masculino como en femenino (Google, 2018).
<i>Incluir etapas de validación y prueba enfocadas en detectar sesgos de género, utilizando métricas específicas que evalúen la justicia y la equidad de los resultados según el género y permitan ajustes antes del despliegue.</i> Enfoque: mitigación de sesgos de género.	Microsoft ha creado un equipo interno dedicado a la ética de la IA, con el objetivo de garantizar que sus productos cumplan con los principios de justicia y equidad. El enfoque de Microsoft incluye la creación de herramientas que permiten a los desarrolladores identificar y corregir sesgos en las fases tempranas del desarrollo de sistemas de IA. Una de estas herramientas es Fairlearn, que evalúa los modelos de IA para asegurar que las predicciones sean justas para todos los grupos sociales, incluidas las mujeres (Raji & Buolamwini, 2019).
<i>Instruir a los equipos de desarrollo y responsables de IA en la identificación y mitigación de sesgos de género en los modelos, promoviendo una cultura de responsabilidad en la creación de IA.</i> Enfoque: identificación, prevención y mitigación de sesgos de género.	Google implementa el programa Machine Learning Fairness para capacitar a sus desarrolladores en la identificación y mitigación de sesgos en IA, incluidos los de género. En este programa, los equipos aprenden a ajustar modelos en productos como Google Photos y Google Translate para evitar resultados estereotipados o discriminatorios. Por ejemplo, en Google Photos, los desarrolladores aplican técnicas para ajustar el reconocimiento de imágenes y evitar la categorización basada en estereotipos de género, y en Google Translate se ofrecen traducciones en ambos géneros en contextos ambiguos para reducir el sesgo (Google, s.f.)
<i>Realizar una evaluación temprana para identificar riesgos de sesgos de</i>	International Business Machines Corporation (IBM) ha desarrollado herramientas como AI Fairness 360 y Finspector

género potenciales en el diseño de los sistemas de IA.

Enfoque: identificación de sesgos de género.

para evaluar y mitigar los sesgos de género en los sistemas de inteligencia artificial. Estas herramientas permiten a los desarrolladores identificar y corregir sesgos en los modelos de IA, promoviendo la equidad y la justicia en sus aplicaciones (IBM, s.f.)

Fuente: DNP-Betagroup, 2024.

En la **etapa de prueba, evaluación, verificación y validación** es importante para garantizar que los sistemas de IA sean justos, transparentes y libres de sesgos de género.

Las buenas prácticas en este caso incluyen identificar y corregir desigualdades, validar tanto el desempeño técnico como la equidad de los modelos, y promover la inclusión de expertos interdisciplinarios y comunidades afectadas. Estas estrategias aseguran sistemas éticos, responsables y respetuosos con la diversidad antes de su despliegue.

Tabla 5 Buenas prácticas en la etapa de prueba, evaluación, verificación y validación

Buena práctica	Ejemplo de aplicación
<i>Realizar auditorías éticas y algorítmicas en las que participen profesionales de diversas disciplinas para analizar los impactos de género en los sistemas de IA, verificando que cumplan con los principios y objetivos de equidad.</i> Enfoque: identificación, prevención y mitigación de sesgos de género.	AI Now Institute, una organización de investigación líder en el análisis de las implicaciones sociales de la IA ha publicado numerosos informes sobre la necesidad de auditar los sistemas de IA para mitigar los sesgos de género. Sus investigaciones han influido en la adopción de buenas prácticas por parte de las empresas tecnológicas, y su enfoque interdisciplinario incluye recomendaciones para asegurar que las auditorías algorítmicas sean obligatorias en las organizaciones que desarrollan IA (Whittaker, y otros, 2018).
<i>Garantizar que los resultados de IA sean comprensibles y que los usuarios puedan cuestionar decisiones o resultados, especialmente aquellos que afectan a grupos de género subrepresentados.</i> Enfoque: mitigación de sesgos de género.	Google aplicó técnicas de explicabilidad en Google Translate para reducir el sesgo de género en sus traducciones automáticas, después de identificar que el sistema tendía a traducir ciertas palabras de manera estereotipada (por ejemplo, "doctor" como masculino y "enfermera" como femenino). Para abordar este problema, Google ajustó su modelo de traducción para que, cuando exista ambigüedad de género, se ofrezcan opciones tanto en masculino como en femenino. Esto no solo permite que los usuarios elijan la traducción adecuada, sino que también fomenta un uso inclusivo del sistema y ayuda a evitar sesgos de género (Google, 2018).
<i>Realizar pruebas que simulen escenarios diversos para asegurar</i>	Una empresa destacada en este campo es Clever AI, una startup que ha diseñado modelos de IA con un enfoque inclusivo. Clever

que el sistema responde de manera equitativa a una variedad de contextos de género.

Enfoque: mitigación de sesgos de género.

Involucrar a personas expertas en estudios de género, activistas y comunidades afectadas en el desarrollo y evaluación de la IA.

Enfoque: identificación, prevención y mitigación de sesgos de género.

Garantizar la transparencia sobre cómo funciona la IA y hacer que los procesos de toma de decisiones sean explicables, para poder identificar y corregir sesgos de género.

Enfoque: identificación y mitigación de sesgos de género.

AI utiliza datos desagregados por género y realiza pruebas en diferentes grupos demográficos para asegurarse de que sus algoritmos funcionen de manera equitativa en diversos contextos (AI4ALL, 2020).

Algorithmic Justice League, ha desempeñado un papel clave en destacar los sesgos raciales y de género en los sistemas de IA. Esta organización trabaja con empresas tecnológicas para ayudarles a identificar y corregir sesgos en sus productos de IA. La Liga también ha sido fundamental en la creación de conciencia pública sobre los riesgos que presentan los algoritmos sesgados y ha impulsado políticas para que las empresas tecnológicas adopten una mayor transparencia en sus procesos de desarrollo. (League, s.f.)

IBM Watson OpenScale permite realizar evaluaciones de equidad en modelos de IA, facilitando la identificación y corrección de sesgos de género, entre otros. Esta herramienta permite a equipos de desarrollo monitorear atributos como "sexo" o "edad" y configurar umbrales de equidad para comparar los resultados entre diferentes grupos, garantizando que los modelos no favorezcan injustamente a un género sobre otro. Además, Watson OpenScale aplica ajustes en tiempo real mediante métodos de atenuación de sesgo, asegurando que los modelos operen de manera justa y transparente mientras están en uso (IBM, 2024).

Fuente: DNP-Betagroup, 2024.

La etapa de **despliegue** es crítica para garantizar que los sistemas de IA mantengan equidad y transparencia al interactuar con los usuarios.

Las buenas prácticas incluyen monitoreo continuo, canales de retroalimentación y ajustes dinámicos según el uso real, asegurando sistemas inclusivos que identifiquen y mitiguen sesgos emergentes. Estas estrategias refuerzan la confianza y promueven la justicia en su aplicación.

Tabla 6 Buenas prácticas en la etapa de entrada en servicio o despliegue

Buena práctica	Ejemplo de aplicación
<i>Implementar un sistema de monitoreo continuo que permita detectar sesgos de género en la operación de IA, especialmente a medida que el sistema interactúa con distintos grupos de usuarios.</i> Enfoque: identificación y mitigación de sesgos de género.	IBM ha sido pionera en la creación de una plataforma de IA ética, que incluye herramientas para mitigar sesgos y promover la transparencia. IBM ha lanzado la plataforma AI Fairness 360, un conjunto de bibliotecas y algoritmos que ayudan a los desarrolladores a identificar, comprender y mitigar los sesgos en los modelos de IA (Bellamy, y otros, 2019). Además, IBM ha implementado políticas de inclusión para asegurar que sus equipos de desarrollo de IA incluyan una representación diversa, lo que ha demostrado ser un factor clave para reducir sesgos en los algoritmos.
<i>Crear canales de comunicación donde las personas usuarias puedan reportar posibles sesgos de género en los resultados, permitiendo ajustes basados en el feedback.</i> Enfoque: identificación y mitigación de sesgos de género.	Google Translate permite a quienes usan esta herramienta reportar errores o posibles sesgos, incluido el sesgo de género, a través de un formulario de comentarios disponible en la parte inferior de la página. Para hacerlo, los usuarios deben acceder a Google Translate, ingresar el texto a traducir, y luego desplazarse hasta la opción "Enviar comentarios" al final de la interfaz. Al hacer clic en esta opción, se abre un formulario donde se pueden detallar los problemas identificados, como el sesgo de género, proporcionando ejemplos específicos (Google Translate, s.f.).

Fuente: DNP-Betagroup, 2024.

La etapa de **explotación y supervisión** es fundamental para mantener la equidad y responsabilidad de los sistemas de IA a lo largo del tiempo. Las buenas prácticas incluyen monitoreo constante, gobernanza ética y documentación de aprendizajes, permitiendo identificar y mitigar sesgos de género, promover mejoras continuas y garantizar una operación justa y transparente.

Tabla 7 Buenas prácticas en la etapa de explotación y supervisión (Monitoreo y ajuste)

Buena práctica	Ejemplo de aplicación
<i>Realizar evaluaciones regulares del impacto de género de los sistemas de IA en uso, revisando cómo afectan a las poblaciones subrepresentadas y ajustando las prácticas según los hallazgos.</i>	El Grupo de Trabajo de Sistemas de Decisión Automatizados de la Ciudad de Nueva York (Ley Local 49 de 2018), tiene como objetivo revisar y recomendar procesos para garantizar que los sistemas de decisión automatizados utilizados por la ciudad sean justos y equitativos. Su labor incluye la realización de evaluaciones periódicas para identificar y mitigar posibles

Enfoque: identificación y mitigación de sesgos de género.	sesgos en los algoritmos empleados, promoviendo así la transparencia y la rendición de cuentas en el uso de la inteligencia artificial en el sector público (Derechos Digitales, 2019).
<i>Establecer un enfoque de mejora continua con roles, métricas e instancias de evaluación y retroalimentación.</i>	IBM ha implementado un enfoque integral para la gobernanza de la IA, estableciendo controles organizacionales a través de su Comité de Ética de IA y un Programa de Gobierno Integrado. Estos mecanismos aseguran que el desarrollo y despliegue de sus sistemas se realicen de manera ética y responsable. Además, IBM ofrece soluciones como Watsonx.governance para ayudar a otras organizaciones a implementar controles organizacionales y técnicos en sus sistemas de IA (World Economic Forum, 2024).
Enfoque: identificación y mitigación de sesgos de género.	
<i>Documentar las lecciones aprendidas en cada fase del ciclo de vida del sistema de IA para promover una mejora continua y facilitar la replicación de buenas prácticas en otros proyectos de IA.</i>	Google Model Cards, desde 2018, Google Research acompaña cada modelo de ML con una “Model Card” que cubre fases de diseño, entrenamiento, evaluación y despliegue, e incluye apartados de “Lecciones aprendidas” donde se registran decisiones clave, errores comunes y recomendaciones para futuros desarrollos (Toolkit, 2020). (UNESCO, 2024)
Enfoque: identificación y mitigación de sesgos de género.	

Fuente: DNP-Betagroup, 2024.

La etapa de **retirada o desmantelamiento** de sistemas de IA es determinante para gestionar de manera ética los impactos de sesgos irreparables, evitando la perpetuación de desigualdades.

Las buenas prácticas incluyen protocolos claros para minimizar daños, proteger a los afectados y reforzar la responsabilidad social, garantizando confianza en el desarrollo de tecnologías futuras.

Tabla 8 Buenas prácticas en la etapa de retirada/desmantelamiento

Buena práctica	Ejemplo de aplicación
En caso de que un sistema de IA presente sesgos irreparables, tener un protocolo de desmantelamiento y un plan de retiro que minimice el daño y evite la perpetuación de dichos sesgos.	Un ejemplo notable de una empresa que implementó un protocolo de desmantelamiento para minimizar el daño y evitar la perpetuación de sesgos es Amazon. En 2018, la compañía desarrolló una herramienta de inteligencia artificial para la selección de personal que, tras su implementación, mostró sesgos de género al favorecer a candidatos masculinos. Al identificar este problema, Amazon decidió

Enfoque: mitigación de sesgos de género.	descontinuar el uso de dicha herramienta para evitar decisiones de contratación injustas y sesgadas (The Verge, 2018).
<i>Evaluar el uso de chatbots de apoyo prenatal basados en IA por parte de mujeres embarazadas al retirar el servicio.</i>	Antes de desactivar un asistente virtual de seguimiento de embarazo, analizar los logs de interacción para identificar consultas frecuentes (p. ej. inquietudes sobre nutrición, síntomas de riesgo), frecuencia de uso por trimestre gestacional y patrones de abandono. Complementar con encuestas breves para determinar qué funcionalidades fueron críticas para la atención de las mujeres y diseñar recursos alternativos (líneas de soporte, apps actualizadas) (Smruti A Mapari 1, 2024).
Enfoque: mitigación de sesgos de género.	

Fuente: DNP-Betagroup, 2024.

Buenas prácticas para la adopción de los sistemas de IA

Incluso cuando los sistemas de IA no son desarrollados en el contexto de aplicación, es importante que se incorporen buenas prácticas en su adopción y uso. El conjunto de orientaciones que se presentan a continuación, son especialmente útiles para entidades u organizaciones que en Colombia están incorporando el uso de sistemas de IA.

Tabla 9 Buenas prácticas para la adopción de los sistemas de IA

Buena práctica	Ejemplo de aplicación
<i>Realizar un análisis exhaustivo de los riesgos éticos, sociales y técnicos, con especial atención a los potenciales sesgos y discriminación que el sistema de IA pueda introducir.</i>	Un ejemplo destacado de esta práctica es el <i>Ethical AI Risk Assessment</i> implementado por Google AI. Google realiza evaluaciones de riesgos éticos y sociales al evaluar los posibles sesgos de género y raza en sus sistemas de IA, a través de auditorías internas y la colaboración con expertos externos para identificar riesgos de discriminación en la adopción de sus sistemas (Google, 2023).
Enfoque: identificación de sesgos de género	
<i>Llevar a cabo un análisis de impacto para anticipar los efectos del sistema de IA sobre los diferentes grupos sociales, considerando aspectos como equidad de género, no discriminación y otros factores éticos.</i>	IBM Watson lleva a cabo estudios de impacto ético en sus sistemas de IA, involucrando a diversos grupos sociales y evaluando los efectos que sus productos pueden tener sobre las personas según su género y grupos poblacionales minoritarios. Esta práctica incluye un análisis continuo de la equidad de género, asegurando que los algoritmos no reproduzcan estereotipos discriminatorios (IBM, 2023).

Enfoque: identificación de sesgos de género

Verificar que el sistema de IA mantenga la explicabilidad, de tal manera que permita comprender cómo y por qué se generan los resultados.

Microsoft ha implementado "Fairlearn", una herramienta que permite a los desarrolladores y usuarios de IA comprender cómo sus modelos toman decisiones, ayudando a identificar cualquier sesgo de género en los resultados y facilitando ajustes para mejorar la transparencia y explicabilidad (Microsoft, 2020).

Enfoque: prevención de sesgos de género

Verificar que la adopción de IA cumpla con las leyes de protección de datos y privacidad, mediante prácticas de anonimización y cifrado, reduciendo el riesgo de exposición indebida.

Apple ha implementado privacidad diferencial en sus sistemas de IA, como Siri, para garantizar la anonimización de los datos de los usuarios, evitando que se recojan datos sensibles relacionados con el género o cualquier otra característica personal, cumpliendo con la normativa de protección de datos como el GDPR (Apple, 2019).

Enfoque: prevención de sesgos de género

Revisar que existan canales de comunicación para que los usuarios puedan expresar preocupaciones o sugerencias, y recibir respuestas sobre cómo se abordarán sus comentarios.

Google permite a los usuarios reportar problemas relacionados con el sesgo de género a través de la opción "Enviar comentarios" en *Google Translate*. Este mecanismo de retroalimentación garantiza que las preocupaciones sobre posibles sesgos sean atendidas, ajustando las traducciones y mejorando la inclusión en el servicio (Google, 2020).

Enfoque: Mitigación de sesgos de género

Fuente: DNP-Betagroup, 2024.

Árbol de decisión

¿Es posible que mi sistema de IA esté reproduciendo desigualdades de género? El Árbol de Decisión que incluye esta guía es una herramienta diseñada precisamente para responder a esa inquietud, de forma clara, guiada y amigable. Esta herramienta no es un diagrama técnico exclusivo para personas expertas: está pensado para cualquier persona o institución que esté desarrollando, usando o evaluando un sistema de IA, desde una aplicación médica hasta un motor de recomendación cultural. Funciona como una brújula que orienta con preguntas sencillas (como en un juego de "sí o no") para que sea posible identificar si hay sesgos de género, en qué parte del ciclo de vida, y qué acciones se pueden tomar para corregirlos o mitigarlos.

Aunque recibe el nombre de “árbol de decisión”, en este caso no se refiere al modelo clásico utilizado en estadística o *machine learning*. Se trata, más bien, de una estructura jerárquica que utiliza la forma de árbol como recurso visual para organizar información. Parte de una idea principal —por ejemplo, la fase del ciclo de vida de un sistema de inteligencia artificial—, se ramifica en ideas secundarias —como los tipos de sesgo— y culmina en elementos específicos, como las acciones. Puede entenderse también como un mapa mental interactivo y navegable, diseñado para facilitar la exploración de rutas posibles según quien lo usa y su contexto. El uso de esta herramienta no requiere instalar ningún software especializado, basta con consultarlo en un navegador web.

Este Árbol de Decisión traduce la compleja taxonomía de sesgos de género en IA en una serie de preguntas binarias y de selección múltiple. Su estructura se fundamenta en los tipos de sesgo (datos, confirmación, interacción, uso, retroalimentación, contextual) definidos en la guía y las fases del ciclo de vida del sistema de inteligencia artificial (planificación, desarrollo, prueba, despliegue, monitoreo y retirada o desmantelamiento). Cada nodo del árbol corresponde a un punto crítico de riesgo identificado en el análisis conceptual de los sesgos

Además, este camino no tiene por qué recorrerse en soledad: al final de cada enunciado, el árbol ofrece una recomendación práctica validada por personas expertas y conectada con herramientas reales que ya están disponibles en línea, lo que permite pasar del diagnóstico a la acción. Por ejemplo:

- ☑ Si eres del equipo de salud y descubres que un modelo de diagnóstico no contempla igual a mujeres y hombres, el árbol puede guiarte hacia la práctica de diversificar tu base de datos y te sugiere herramientas específicas para hacerlo.
- ☑ Si trabajas en el sector financiero y te das cuenta de que tu IA tiende a reproducir sesgos históricos en el análisis crediticio, el árbol te orienta para implementar

monitoreo continuo con métricas de equidad y te enlaza con recursos como AI Fairness 360.

Esta herramienta representa un cambio en cómo abordamos los desafíos éticos de la tecnología. No se trata solo de corregir errores, sino de construir sistemas más justos, representativos y confiables desde el principio. Es una invitación a prevenir, actuar, aprender y transformar en nuestra interacción con la IA.

Su estructura se fundamenta en los tipos de sesgo (datos, confirmación, interacción, uso, retroalimentación, contextual) definidos en la guía y las fases del ciclo de vida del sistema de inteligencia artificial (planificación, desarrollo, prueba, despliegue, monitoreo y retirada/desmantelamiento). Cada nodo del árbol corresponde a un punto crítico de riesgo identificado en el análisis conceptual de los sesgos y se construye en consonancia con referentes para la detección, prevención y mitigación de sesgos.

¿Cómo se hizo?

La construcción del Árbol de Decisión fue un proceso colaborativo y participativo que buscó integrar diversas voces del ecosistema digital. Se combinó el conocimiento técnico con experiencias reales de quienes enfrentan, regulan o investigan los sesgos de género en la aplicación de inteligencia artificial⁶. Esto permite traducir conceptos complejos en rutas comprensibles, visuales y navegables para cualquier perfil de uso. Además, se priorizó la accesibilidad para garantizar que cualquier persona (sin importar su nivel técnico) pueda utilizar la herramienta con confianza.

La herramienta fue diseñada para que su uso no requiera formación técnica especializada: basta con contar con un computador o dispositivo con acceso a internet y conocimientos básicos de navegación web. Esta combinación interdisciplinaria, sumada a su enfoque centrado en la usabilidad, hace del Árbol de Decisión una herramienta fiable, ética y accionable frente a los desafíos de la inteligencia artificial con perspectiva de género.

El paso a paso de este proceso consistió en:

- ☑ **Extracción de criterios:** se definieron actores de interés como personas desarrolladoras, reguladoras, académicas, ONGs, usuarias y medios de comunicación y se utilizó una matriz de influencia-interés para definir los nodos de decisión clave y la prioridad que debía abordarse para cada actor y contexto. Con estos actores se realizaron entrevistas, grupos focales y talleres de *design-thinking*, cuyos insumos fueron procesados y codificados por temas, lo que permitió identificar los principales comportamientos de riesgo y barreras de género en cada fase de la IA.

⁶ Se emplearon diez (10) instrumentos de recolección cualitativa: entrevistas semiestructuradas y en profundidad, grupos focales, talleres participativos, foros virtuales, paneles de personas expertas, mesas de trabajo, cuestionarios de preguntas abiertas, análisis de redes sociales (ARS), y *web scraping* mediante API. Estos permitieron captar experiencias, percepciones y datos clave de diversos actores del ecosistema, garantizando un insumo robusto para la construcción y validación del árbol de decisión.

- ☑ **Construcción técnica:** el diseño técnico del árbol de decisión parte de una estructura jerárquica tipo mapa mental, que permite representar relaciones entre fases del ciclo de vida, tipos de sesgo y actores responsables. Esta estructura se codificó utilizando formatos ligeros, lo que permite que el árbol funcione sin necesidad de conexión a bases de datos complejas ni instalación de software especializado. También se aplicaron principios de usabilidad y accesibilidad (WCAG 2.1) para garantizar compatibilidad en navegadores móviles y de escritorio sin necesidad de instalar software adicional. La codificación utilizada facilitó la navegación lógica e interactiva y también garantizó que la herramienta pueda crecer en el futuro, permitiendo añadir nuevas fases, tipos de sesgo o actores según las necesidades de actualización o contexto institucional.
- ☑ **Validación:** finalmente, se realizaron pruebas piloto para ajustar términos y opciones de respuesta. La validación incorporó equipos interdisciplinarios, con personas expertas en género, IA y política pública. También se revisó su alineación con las directrices del CONPES 4080 y estándares internacionales como los de la UNESCO y la OCDE. Este enfoque garantizó que el árbol no solo funcione como guía técnica, sino también como un dispositivo pedagógico y de transformación institucional, capaz de promover sistemas de IA más inclusivos, justos y representativos.

¿Para qué sirve?

El Árbol de Decisión es una herramienta diseñada para generar conciencia, corresponsabilidad y acción concreta frente a los sesgos de género en sistemas de inteligencia artificial. Facilita la conversación entre áreas técnicas, administrativas y sociales, permitiendo que equipos multidisciplinarios tomen decisiones informadas desde una perspectiva de género.

También funciona como insumo pedagógico para procesos de formación, auditoría o diseño de políticas públicas en entornos digitales. Su estructura accesible permite que personas con distintos roles y niveles de conocimiento técnico identifiquen riesgos y sepan cómo actuar frente a ellos. Puntualmente, esta herramienta le sirve a las personas usuarias y a Colombia para:

- ☑ Cerrar la brecha entre diagnóstico y acción, al ofrecer rutas claras desde la detección del sesgo hasta la implementación de soluciones.
- ☑ Fortalecer la gobernanza ética de la IA, alineando decisiones con principios de equidad, estándares internacionales (OCDE, UNESCO) y el CONPES 4080.
- ☑ Impulsar la transformación digital inclusiva, asegurando que el desarrollo de inteligencia artificial incorpore principios de justicia social y equidad de género.
- ☑ Identificar y caracterizar sesgos de género presentes en un proyecto de IA, contextualizados según la fase del ciclo de vida y el rol de quien lo aborda, lo que orienta también la definición de responsabilidades.

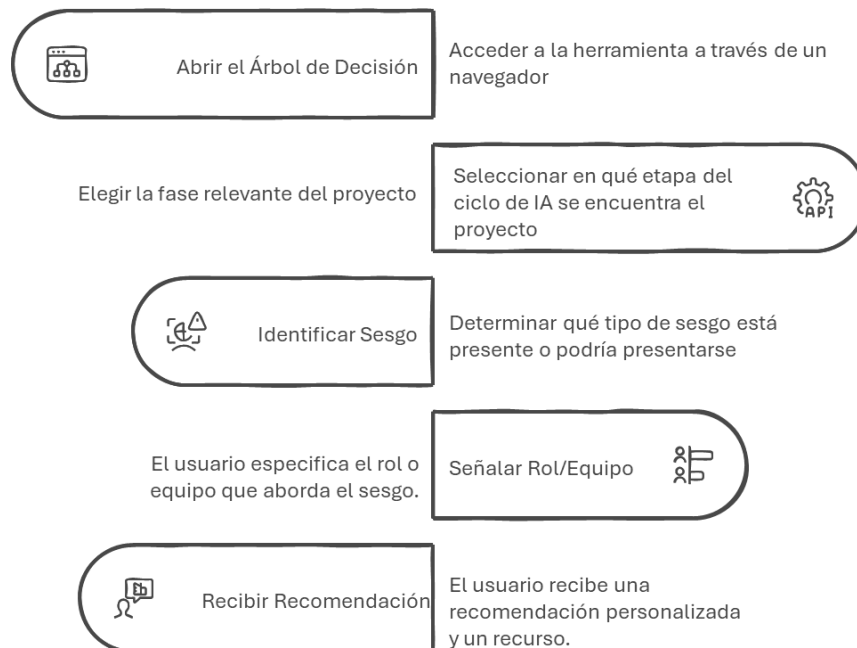
- ✓ Priorizar intervenciones y acceder directamente a soluciones concretas, pues el árbol ofrece al final de cada ruta un enlace que lleva directamente a unas herramientas. Esto permite avanzar desde la identificación del sesgo, hasta la toma de una decisión informada y aplicar una acción específica con respaldo técnico.

¿Cómo se usa?

Usar el Árbol de Decisión es muy sencillo e intuitivo. Solo es necesario ingresar a: <https://ucd-dnp.github.io/sesgosIA/page>

Una vez allí, se siguen las instrucciones en pantalla: se lee cada pregunta, se hace clic sobre la opción que más se ajuste a un caso en concreto y a continuación se navega por las rutas que se irán desplegando automáticamente. Al final, se recibe una recomendación práctica junto con el enlace directo al recurso que se puede utilizar. Si existe duda sobre el funcionamiento o se quiere ver un ejemplo guiado paso a paso, también es posible consultar el video tutorial disponible en:

Figura 2 Ruta de navegación del árbol de decisión



Fuente: elaboración propia con ayuda de Napkin

Es posible recorrerlo las veces que se necesite, explorar diferentes rutas y descargar las buenas prácticas sugeridas como insumo para el equipo o institución. Es posible categorizar y gestionar riesgos asociados con sesgos de género en las soluciones de IA, donde

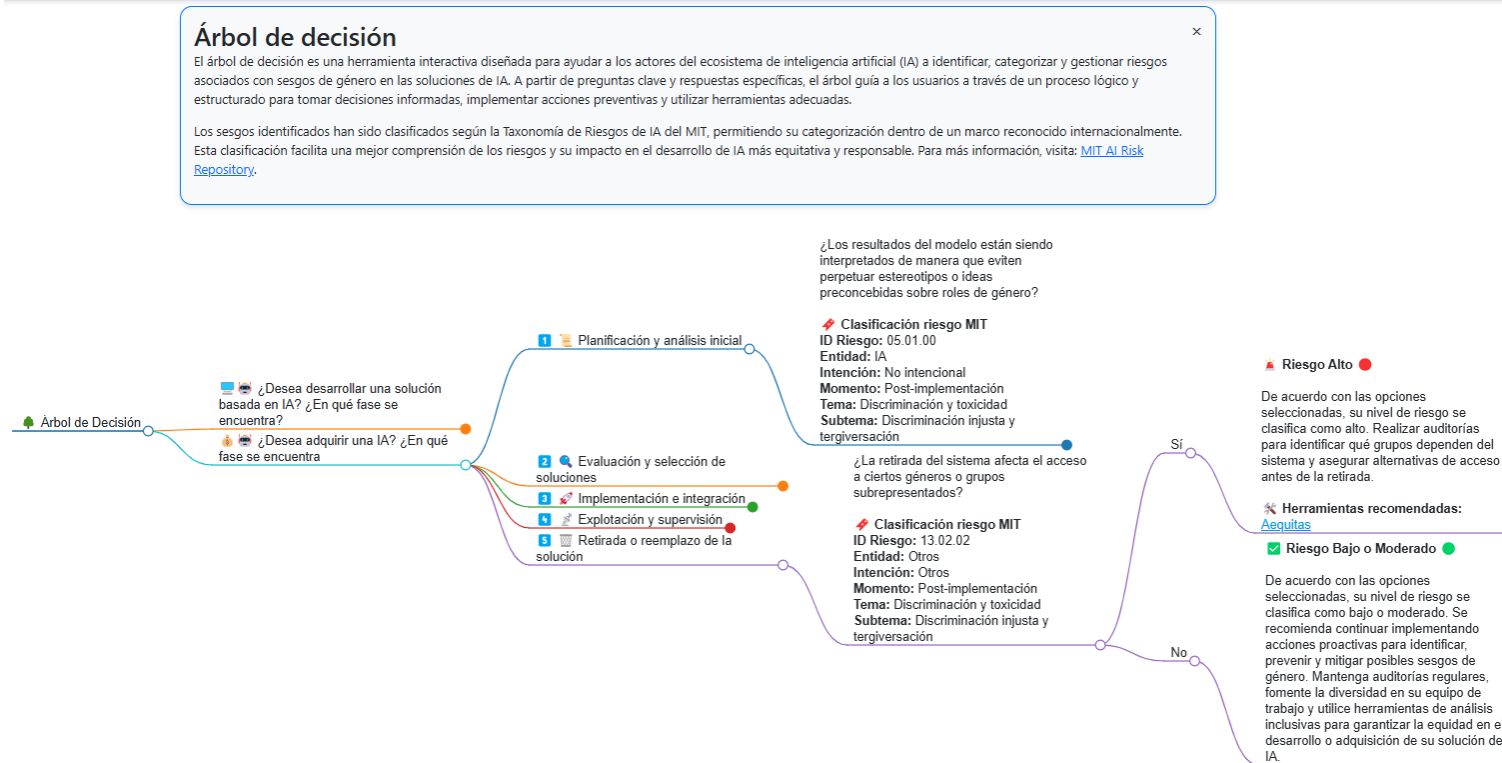


a partir de preguntas clave y respuestas específicas, el árbol guía a los usuarios a través de un proceso lógico y estructurado para tomar decisiones informadas, implementar acciones preventivas y utilizar herramientas adecuadas.

El árbol de decisión es mucho más que una guía interactiva, es una herramienta estratégica que traduce conceptos técnicos complejos en rutas simples y accesibles para toda persona interesada en construir soluciones de inteligencia artificial justas, responsables y con enfoque de género.

La siguiente imagen ilustra su estructura lógica y navegable, incentivando su consulta como una brújula para mejorar procesos y proyectos de Inteligencia Artificial con perspectiva de equidad. Se trata de un recurso que puede ser usado tantas veces como sea necesario, en cualquier fase del ciclo de vida de la inteligencia artificial, promoviendo la mejora continua y el compromiso con la ética digital.

Figura 3 Representación gráfica del árbol de decisión: navegación guiada e interactiva



Fuente. Elaboración propia, recuperado de: <https://ucd-dnp.github.io/sesgosIA/page>

Catálogo de herramientas

¿Qué es?

El Catálogo de Herramientas es un recurso práctico y accesible que acompaña al árbol de decisión para pasar del diagnóstico a la acción. Recoge una selección de librerías, guías, listas de verificación y metodologías (tanto de código abierto como privativos) que permiten identificar, prevenir o mitigar los sesgos de género a lo largo del ciclo de vida de los sistemas de inteligencia artificial.

Este catálogo está pensado para que cualquier actor del ecosistema (e.g.: personas desarrolladoras, tomadoras de decisión, equipos de auditoría, responsables de compras, organizaciones sociales y educativas) pueda encontrar herramientas confiables y contextualizadas, adaptadas a sus necesidades y niveles de experiencia.

Cada herramienta (librerías, *frameworks*, guías y *check-lists*) está organizada en fichas breves y claras que incluyen:

- ✳ El nombre y versión de la herramienta.
- ✳ Su autoría y requisitos técnicos mínimos.
- ✳ La fase del ciclo de vida donde puede aplicarse.
- ✳ El tipo de sesgo que ayuda a mitigar (por ejemplo, de datos, uso o retroalimentación).
- ✳ Su objetivo principal: identificar, prevenir o mitigar.
- ✳ Una guía paso a paso para su uso y un enlace directo para acceder a ella.

Gracias a esta estructura, el catálogo se convierte en una herramienta práctica para la toma de decisiones: identifica soluciones potenciales y permite compararlas según criterios como facilidad de implementación, tipo de licencia o cobertura técnica.

¿Cómo se hizo?

El Catálogo de Herramientas fue construido mediante un proceso riguroso y colaborativo que combinó revisión técnica, participación multisectorial y diseño centrado en la usabilidad. Su objetivo fue garantizar que cada herramienta incluida sea pertinente, accesible y aplicable al contexto colombiano, con un enfoque claro en la equidad de género. Este proceso se puede resumir en tres etapas principales:

- ☑ **Búsqueda y selección de herramientas:** se realizó una revisión sistemática de literatura académica reconocida (como IEEE y ACM), análisis de APIs especializadas en equidad algorítmica (como *AI Fairness 360* y *Fairlearn*), y consulta de repositorios públicos como GitHub. Esta exploración permitió identificar recursos técnicamente sólidos, relevantes y de uso extendido a nivel internacional.
- ☑ **Validación participativa:** los recursos identificados fueron discutidos y evaluados en talleres participativos con seis grupos de actores clave del ecosistema digital:



personas desarrolladoras, reguladoras, academia, ONGs, medios y personas usuarias. Esta validación permitió asegurar que las herramientas seleccionadas fueran útiles, comprensibles y alineadas con los desafíos reales que enfrentan distintos perfiles institucionales.

- ☑ **Estructuración de la ficha:** Cada recurso fue organizado en una ficha técnica siguiendo los principios de los datos FAIR (que sean encontrables, accesibles, interoperables y reutilizables). Además, se incluyeron campos adicionales como las métricas de impacto, facilitando el seguimiento y evaluación de su uso a lo largo del tiempo.

¿Para qué sirve?

El Catálogo de Herramientas está diseñado para funcionar como un kit práctico y confiable al servicio de quienes trabajan con inteligencia artificial desde una perspectiva ética y con enfoque de género. Su principal propósito es facilitar la acción concreta, brindando acceso inmediato a recursos que han sido previamente seleccionados, validados y organizados para apoyar decisiones informadas.

Esta herramienta permite:

- ☑ Acceder fácilmente a librerías y guías que aplican las buenas prácticas sin necesidad de búsquedas adicionales.
- ☑ Estandarizar procesos de mitigación de sesgos y adoptar protocolos comunes (como auditorías, documentación de conjuntos de datos o configuraciones éticas), promoviendo consistencia en la implementación de buenas prácticas en IA.
- ☑ Comparar herramientas según criterios de licencia, facilidad de integración y cobertura del tipo de sesgo.

En conjunto, el catálogo no solo entrega soluciones técnicas, sino que impulsa una cultura institucional orientada a la equidad, la transparencia y el aprendizaje continuo en entornos digitales. Es, en esencia, una guía viva para construir tecnologías más justas, inclusivas y responsables.

¿Cómo se usa?

Explorar el Catálogo de Herramientas es fácil, gratuito y accesible para todas las personas. Solo necesitas ingresar a: <https://ucd-dnp.github.io/sesgosIA/page>

Desde cualquier navegador (en computador o dispositivo móvil) podrás filtrar las herramientas según tu rol, la fase del proyecto de IA, el tipo de recurso que buscas y el objetivo (identificar, prevenir o mitigar sesgos). La plataforma está diseñada para que cualquier persona, sin necesidad de conocimientos técnicos avanzados, pueda encontrar recursos útiles y confiables.

Cada ficha incluye un resumen ejecutivo, pasos para la implementación y un enlace directo al recurso original, lo que te permite pasar rápidamente del diagnóstico a la acción. Si tienes dudas sobre cómo usarlo, muy pronto estará disponible un video tutorial para acompañarte paso a paso. A continuación, se describen los pasos para su uso:

1. Acceso

Ingresar al catálogo mediante el siguiente enlace: <https://ucd-dnp.github.io/sesgosIA/page>

El sitio está optimizado para navegadores comunes y funciona sin necesidad de crear una cuenta.

2. Filtrado

Una vez dentro, es posible personalizar la búsqueda usando los siguientes filtros:

- ✧ **Actor:** Se puede seleccionar si es desarrollador/a, regulador/a, auditor/a, usuario/a, etc.
- ✧ **Fase del ciclo de vida:** Desde la planificación hasta el desmantelamiento del sistema de IA.
- ✧ **Objetivo:** Identificar, prevenir o mitigar el sesgo de género.
- ✧ **Tipo de recurso:** Es posible elegir entre librerías, guías, listas de verificación (checklists) y marcos metodológicos.
- ✧ Estos filtros permiten adaptar la búsqueda a las necesidades específicas de cada persona o entidad, asegurando que los recursos sugeridos sean útiles y contextualizados.

3. Exploración de ficha

Al seleccionar una herramienta, se despliega una ficha que contiene:

- ✧ Un resumen ejecutivo que explica de forma simple para qué sirve.
- ✧ Los pasos para su implementación o configuración.
- ✧ Información sobre su licencia, accesibilidad técnica y tipo de sesgo que aborda.
- ✧ Un enlace directo al recurso original (repositorio, sitio web o documento técnico).

Este catálogo, además de ser un repositorio, es un puente entre la detección del problema y la solución. Su diseño responde a las necesidades levantadas en entrevistas, talleres y paneles realizados con actores clave del ecosistema colombiano de IA, garantizando su pertinencia, aplicabilidad y facilidad de uso.

El catálogo de herramientas es una extensión práctica de esta guía, su valor no radica únicamente en listar recursos, sino en organizar y presentar de manera intuitiva soluciones confiables que pueden ser consultadas en línea de forma sencilla y rápida.

Su diseño visual permite identificar de un vistazo los elementos clave de cada recurso, como su objetivo, licencia y aplicabilidad y navegarlo de forma personalizada gracias a filtros

dinámicos. Esta estructura no solo apoya la toma de decisiones informadas, sino que fomenta la exploración activa y el aprendizaje continuo.

La siguiente imagen ilustra cómo se ve y cómo facilita a cualquier persona (sin importar su nivel técnico) explorar las herramientas sugeridas para identificar, prevenir o mitigar sesgos de género en inteligencia artificial.

Figura 4 Catálogo de Herramientas: estructura y filtros para explorar recursos



Fuente. Elaboración propia, recuperado de: <https://ucd-dnp.github.io/sesgosIA/page>

Esta guía busca ser un recurso práctico para que cualquier persona u organización que participe en el ciclo de vida de la inteligencia artificial pueda identificar, prevenir y mitigar sesgos de género de manera efectiva. Su valor radica en ofrecer orientaciones y herramientas que facilitan la toma de decisiones informadas y responsables, contribuyendo a que el desarrollo y la adopción de la IA en Colombia se hagan con criterios de equidad y respeto por la diversidad. La apropiación de este documento por parte de todos los actores es clave para avanzar hacia un ecosistema digital más justo y representativo.

Referencias

- Accenture. (s.f.). *Trabajar para acelerar la igualdad para todos*. Obtenido de <https://www.accenture.com/co-es/about/inclusion-diversity/gender-equality#:~:text=Hemos%20establecido%20metas%20ambiciosas%20para,para%20personas%20con%20g%C3%A9nero%20binario.>
- AI4ALL. (2020). *Building the Future of AI: Diverse Talent, Inclusive AI Education*. Obtenido de <https://ai-4-all.org/>
- Alta Consejería Distrital de TIC. (2023). *Inteligencia artificial en Bogotá-Región*. Alcaldía Mayor de Bogotá - Agata - Invest in Bogotá. Obtenido de <https://en.investinbogota.org/wp-content/uploads/2023/08/BEN-31-AI.pdf>
- Álvarez, L., & Rizzi, A. (2023). *Una guía de gestión de riesgos para inversores de impacto*. Center for Financial Inclusion - USAID. Obtenido de <https://www.centerforfinancialinclusion.org/wp-content/uploads/2024/03/Espanol-Equitable-AI-Toolkit.pdf>
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (23 de Mayo de 2016). *Machine Bias*. Recuperado el 18 de 09 de 2024, de <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Apple. (28 de Agosto de 2019). *Mejoras en las protecciones de privacidad de Siri*. Obtenido de <https://www.apple.com/co/newsroom/2019/08/improving-siris-privacy-protections/>
- Apple. (s.f.). *Privacy. That's Apple*. Obtenido de <https://www.apple.com/privacy/>
- Bellamy, R., Dey, K., Hind, M., Hoffman, S., Houde, S., Kannan, K., . . . Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *IBM Journal of Research and Development*, 63(4/5), 63(4/5), 4-1.
- Bolukbasi, T., Chang, K., Zou, J. Y., & Saligrama, V. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *30th International Conference on Neural Information Processing Systems*, (págs. 4349–4357). doi:<https://doi.org/10.48550/arXiv.1607.06520>
- Buolamwini, J., & Gebru, T. (2018). When women or minorities are underrepresented in datasets, algorithms are likely to perform poorly for these groups, reinforcing existing inequalities. *Proceedings of Machine Learning Research*, (págs. 1-15).
- Business & Marketing School. (Agosto de 2024). *¿Qué es un sesgo algorítmico y cómo se produce? Ejemplos*. Obtenido de <https://www.esic.edu/rethink/tecnologia/que-es-un-sesgo-algorimico-y-como-se-produce-ejemplos-c>

- C3.AI. (s.f.). *IA explicable ¿qué es la IA explicable?* Obtenido de <https://c3.ai/glossary/machine-learning/explainability/>
- C4IR.CO. (s.f.). *IA Inclusiva Herramientas para el uso responsable de la Inteligencia Artificial*. Obtenido de <https://c4ir.co/un-kit-para-identificar-sesgos-de-genero-en-inteligencia-artificial/>
- CENIA. (2024). *Índice Latinoamericano de Inteligencia Artificial*. Obtenido de https://indicelatam.cl/wp-content/uploads/2025/01/ILIA_2024_020125_compressed.pdf
- Consejo Nacional para Prevenir la Discriminación CONAPRED. (s.f.). *¿Qué es la discriminación?* Obtenido de <https://www.conapred.org.mx/discriminacion-en-mexico/que-es-la-discriminacion/>
- Corte Constitucional de Colombia. (2024). *Sentencia T-323 DE 2024* . Obtenido de DERECHO AL DEBIDO PROCESO-Criterios orientadores para el adecuado uso de la inteligencia artificial (IA): <https://www.corteconstitucional.gov.co/relatoria/2024/t-323-24.htm>
- De Mauro, A., Greco, M., & Grimaldi, M. (4 de Abril de 2016). *A formal definition of Big Data based on its essential features*. Obtenido de <https://www.emerald.com/insight/content/doi/10.1108/LR-06-2015-0061/full/html>
- Derechos Digitales. (2019). *Políticas públicas e Inteligencia Artificial. Una lista de buenas prácticas para tener en consideración*. Obtenido de https://ia.derechosdigitales.org/wp-content/uploads/2022/05/DD_IA_03.pdf
- DNP. (2022). *Documento CONPES 4080: Política Pública de Equidad de Género para las Mujeres*. Obtenido de <https://colaboracion.dnp.gov.co/cdt/Conpes/Econ%C3%B3micos/4080.pdf>
- DNP. (2023). *Plan Nacional de Desarrollo (PND) 2022-2026 "Colombia Potencia Mundial de la Vida"*. Obtenido de <https://colaboracion.dnp.gov.co/CDT/portalDNP/PND-2023/2023-05-04-bases-plan-nacional-de-inversiones-2022-2026.pdf>
- DNP. (2024). *Estrategia Nacional Digital (END) 2023-2026*. Obtenido de https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Digital/EVENTOS/END_Colombia_2023_2026.pdf
- DNP. (2025). *CONPES 4144: Política Nacional de Inteligencia Artificial* . Obtenido de <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/4144.pdf>
- Doshi, F., & Kim, B. (2 de Marzo de 2017). *Towards a rigorous science of interpretable machine learning*. Obtenido de <https://arxiv.org/abs/1702.08608>
- Exponiendo.ai. (2019). *La diversidad en los rostros de IBM*. Obtenido de https://exposing.ai/ibm_dif/

- Fondo Monetario Internacional. (Diciembre de 2022). *La brecha digital de género*. Obtenido de <https://www.imf.org/es/Publications/fandd/issues/2022/12/the-digital-gender-gap-khera-ogawa-sahay-vasishth>
- FRA. (2022). *Bias in Algorithms. Artificial Intelligence and Discrimination*. Vienna: European Union Agency For Fundamental Rights. Obtenido de https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf
- GLAAD. (24 de Junio de 2021). *GLAAD se asocia con Google para ofrecer recursos de marketing LGBTQ “All In”*. Obtenido de <https://glaad.org/announcing-glaad-partnership-google-all/>
- Gobierno Nacional de Colombia. (2021). *Marco Ético para la Inteligencia Artificial en Colombia*. Obtenido de <https://minciencias.gov.co/sites/default/files/marco-etico-ia-colombia-2021.pdf>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, Massachusetts: The MIT Press.
- Google. (6 de Diciembre de 2018). *Cómo reducir el sesgo de género en Google Translate*. Obtenido de <https://blog.google/products/translate/reducing-gender-bias-google-translate/>
- Google. (17 de Marzo de 2020). *Ayudas de Google Translate*. Obtenido de https://support.google.com/translate/answer/12111375?hl=es-ES&utm_source
- Google. (Octubre de 2023). *Evaluating social and ethical risks from generative AI*. Obtenido de <https://deepmind.google/discover/blog/evaluating-social-and-ethical-risks-from-generative-ai/>
- Google AI. (2020). *Building fairness into AI*. Obtenido de <https://ai.google/research>
- Google. (s.f.). *Conceptos del AA*. . Obtenido de <https://developers.google.com/machine-learning/crash-course/fairness?hl=es-419>
- Google. (s.f.). *What-If Tool*. Obtenido de <https://pair-code.github.io/what-if-tool/>
- Google Translate. (s.f.). *Google Translate*. Obtenido de <https://translate.google.com/?hl=es&sl=en&tl=es&op=translate>
- IBM. (2023). *Informe IBM Impact de 2023*. Obtenido de <https://www.ibm.com/downloads/documents/us-en/10a99803d62fd3a5>
- IBM. (25 de Octubre de 2024). *Evaluaciones de equidad*. Obtenido de <https://datapatform.cloud.ibm.com/docs/content/wsj/model/wos-monitor-fairness.html?context=cpdaas&locale=es>
- IBM. (s.f.). *AI Fairness 360*. Obtenido de <https://aif360.res.ibm.com/>

- Instituto Carl Rogers. (7 de Junio de 2023). *¿Qué es la perspectiva de género y por qué es importante?* Obtenido de <https://www.institutocarlgrogers.org/perspectiva-de-genero/>
- Kossow, N., Svea, W., Svea, W., & Jenkins, M. (05 de Febrero de 2021). *Algorithmic transparency and accountability*. Obtenido de https://knowledgehub.transparency.org/assets/uploads/kproducts/Algorithmic-Transparency_2021.pdf
- Latif, E., Zhai, X., & Liu, L. (26 de Diciembre de 2023). *AI Gender Bias, Disparities, and Fairness: Does Training Data Matter?* *arXiv*. Obtenido de <https://arxiv.org/pdf/2312.10833v2>
- League, A. J. (s.f.). Obtenido de https://www-ajl-org.translate.goog/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es-419&_x_tr_pto=sc
- Microsoft. (19 de Mayo de 2020). *AI The Show*. Obtenido de <https://learn.microsoft.com/es-es/shows/ai-show/building-fairer-ai-systems-with-fairlearn>
- Ministerio de Ciencia, Tecnología e Innovación. (2024). *Hoja de Ruta para el Desarrollo y Aplicación de la Inteligencia Artificial en Colombia*. Ministerio de Ciencia, Tecnología e Innovación. Obtenido de https://minciencias.gov.co/sites/default/files/upload/noticias/hoja_de_ruta_adopcion_etica_y_sostenible_de_inteligencia_artificial_colombia_0.pdf
- Mójica Muñoz, K. S. (2022). *Retos de la inteligencia artificial en la solución de problemas públicos en Colombia*. Oportunidad Estratégica. Obtenido de <https://oportunidadestrategica.com/wp-content/uploads/2023/05/RETOS-DE-LA-INTELIGENCIA-ARTIFICIAL-EN-LA-SOLUCION-DE-PROBLEMAS-PUBLICOS-EN-COLOMBIA.pdf>
- National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Obtenido de <https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf#:~:text=Bias%20is%20broader%20than%20demographic,considered%20and%20managed%3A%20systemic%2C%20computational>
- NetHope's Center for the Digital Nonprofit. (2021). *Kit de Herramientas de IA Equitativa (NetHope)*. Obtenido de file:///C:/Users/ypinto/Downloads/NetHope_Gender%20Equitable%20AI%20Toolkit_240322.pdf
- OCDE. (2019). *Recomendaciones sobre Inteligencia Artificial*. Obtenido de <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- OCDE. (2 de Mayo de 2024). *Recommendation of the Council on Artificial Intelligence*. Obtenido de <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>

- OECD. (2024). *Governing with Artificial Intelligence: Are Governments Ready?* Obtenido de https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/governing-with-artificial-intelligence_f0e316f5/26324bc2-en.pdf?utm_source
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.
- ONU Mujeres. (5 de Febrero de 2025). Obtenido de Cómo la inteligencia artificial refuerza los sesgos de género y qué podemos hacer al respecto : <https://www.unwomen.org/es/noticias/entrevista/2025/02/como-la-inteligencia-artificial-refuerza-los-sesgos-de-genero-y-que-podemos-hacer-al-respecto#:~:text=Esta%20situaci%C3%B3n%20se%20denomina%20sesgos,sesgado s%20con%20que%20se%20aliment%C3%B3>.
- Organización Mundial de la Salud (OMS). (23 de Agosto de 2018). *Genero y salud*. Recuperado el 20 de 09 de 2024, de <https://www.who.int/es/news-room/fact-sheets/detail/gender>
- Oxford Insights. (2024). *Government AI Readiness Index* . Obtenido de <https://oxfordinsights.com/ai-readiness/ai-readiness-index/>
- Raji, I., & Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (págs. 429-435). New York: Association for Computing Machinery.
- Real Academia Española RAE. (s.f.). *Definición de algoritmo*. Obtenido de <https://dle.rae.es/algoritmo>
- Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). California: Pearson.
- Stuart J, R., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (Vol. 4th edition). Pearson .
- Suchman, L. (2002). Located Accountabilities in Technology Production. *Scandinavian Journal of Information Systems*, 14(2). Obtenido de <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1082&context=sjis>
- The Verge. (10 de Octubre de 2018). *Amazon descarta una herramienta de reclutamiento de inteligencia artificial interna que estaba sesgada contra las mujeres*. Obtenido de <https://www.theverge.com/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report>
- UNESCO. (Noviembre de 2021). *Recomendación sobre la ética de la inteligencia artificial*. Obtenido de <https://unesdoc.unesco.org/ark:/48223/pf0000380455.page>

- UNESCO. (Noviembre de 2021). *Recomendación sobre la ética de la inteligencia artificial*.
Obtenido de <https://www.unesco.org/es/articles/recomendacion-sobre-la-etica-de-la-inteligencia-artificial>
- UNESCO. (2021). *Recomendación sobre la Ética de la Inteligencia Artificial*. Obtenido de <https://unesdoc.unesco.org/ark:/48223/pf0000380455>
- UNESCO. (2024). *Challenging Systematic Prejudices: An Investigation into Bias Against Women and Girls in Large Language Models*. París: UNESCO.
- UNESCO. (Febrero de 2024). *Ética de la inteligencia artificial. La recomendación*. Obtenido de <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>
- UNESCO. (2024). *Women for Ethical AI. Outlook Study on Artificial Intelligence and Gender*. París: UNESCO .
- Unesco. (2025). *Tackling Gender Bias and Harms in Artificial Intelligence (AI)*. Obtenido de <https://www.unesco.org/en/articles/tackling-gender-bias-and-harms-artificial-intelligence-ai#:~:text=As%20generative%20AI%20,those%20impacting%20women%20and%20girls>
- UNESCO. (s.f.). *Inteligencia Artificial*. Obtenido de <https://www.unesco.org/es/artificial-intelligence#:~:text=Construida%20a%20partir%20de%20data,ling%C3%BC%C3%A>
Dstica%20y%20hasta%20la%20creatividad.
- UNESCO, IRCAI. (2024). *Challenging systematic prejudices: an Investigation into Gender Bias in Large Language Models*. Obtenido de <https://unesdoc.unesco.org/ark:/48223/pf0000388971>
- UNFPA. (s.f.). *Preventing Technology-Facilitated Gender-Based Violence (TF GBV)*. Obtenido de https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/GDC-Submission_UNFPA.pdf
- Viteri, M. A., & Gómez Pineda, M. I. (29 de Noviembre de 2023). Obtenido de Blog BID.
Hablemos de Sostenibilidad y Cambio Climático:
<https://blogs.iadb.org/sostenibilidad/es/inteligencia-artificial-y-equidad-de-genero-un-espejo-de-nuestras-sociedades/>
- Viveros Vigoya, M. (2023). *Interseccionalidad, giro decolonial y comunitario*. CLACSO-TNI.
Obtenido de <https://biblioteca-repositorio.clacso.edu.ar/bitstream/CLACSO/248817/1/Interseccionalidad.pdf>
- Wajcman, J. (2006). *El Tecnofeminismo*. Ediciones Cátedra.
- Whittaker, M., Crawford, K., Dobbe, R., Fried , G., Kaziunas, E., Mathur, V., & Schwartz, O. (6 de Diciembre de 2018). *AI now report 2018*. Obtenido de <https://ainowinstitute.org/publication/ai-now-2018-report-2>

World Economic Forum. (20 de Septiembre de 2024). *Gobernanza de la IA: Cómo la regulación, la colaboración y la demanda de habilidades están configurando las tendencias*. Obtenido de <https://es.weforum.org/stories/2024/09/gobernanza-de-la-ia-como-la-regulacion-la-colaboracion-y-la-demanda-de-talento-estan-dando-forma-a-la-industria/>

Ziwei, L., Ping, L., Xiaogang, W., & Xiaoou, T. (24 de Septiembre de 2015). *Large-scale CelebFaces Attributes (CelebA) Dataset. Deep Learning Face Attributes in the Wild*. Obtenido de <https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>