Fecha de recepción: febrero 2026
Fecha de aprobación: marzo 2026

Inteligencia Artificial en la Industria de la Moda Discriminación y Justicia algorítmica

Silvana Pampuri(*)

Resumen: La incorporación de sistemas de inteligencia artificial (IA) en la industria del diseño y de la moda, transforma de manera profunda la cadena de valor, desde la ideación y el diseño hasta la producción, el marketing, la comercialización, la logística y la experiencia posventa. Su impacto es cada vez más profundo, efectuando aportes estratégicos y también presentando desafíos éticos, técnicos y legales convergiendo la sostenibilidad, equidad y justicia.
Estos sistemas no son neutros, ni en sus ventajas ni en su riesgos, ya que por la velocidad de su propagación pueden reproducir y amplificar sesgos históricos, sociales e institucionales. Hace tiempo que venimos escuchando de justicia y equidad "algorítmica".
Para ello se exploran conceptos y clasificaciones de sesgos algoritmos, medios de mitigación desde una perspectiva sociotécnica y se abordan principios éticos, fairness y greenwashing algorítmico.

Palabras clave: inteligencia artificial - sesgos algorítmicos - justicia algorítmica - discriminación algorítmica - greenwhashing algorítmico.

[Resúmenes en inglés y portugués en la página 251]

(*) Abogada, egresada de la Universidad de Buenos Aires, Mediadora, Corredora Pública y Titulada en Escribanía. En ejercicio liberal de la profesión de abogada desde 1994. Integrante del Instituto del Derecho de la Moda del Colegio de la Abogacía de la Ciudad Autónoma de Buenos Aires. Integrante titular por concurso de la ex Clínica Jurídica de los Derechos de los Niños, Niñas y Adolescentes, en convenio entre UNICEF y CALZ. Ex jefa de trabajos prácticos de Derecho Constitucional y Derecho Público, Cátedra Dr. Eduardo Rubén Florio. Email de contacto: silvana@silvanapampuri.com

1. Introducción

En los últimos años la industria de la moda se ha convertido en un escenario privilegiado para la adopción de sistemas de inteligencia artificial.
Plataformas de análisis de tendencias procesan datos provenientes de redes sociales, búsquedas en internet, interacciones en sitios de comercio electrónico y registros de ventas para anticipar preferencias de consumidores y decidir qué producir, cuándo y en qué cantidades. Herramientas de inteligencia artificial generativa asisten en la creación de estampas, diseños, prototipos digitales y campañas visuales con modelos sintéticos. Sistemas de optimización logística y de gestión de inventario integran información de proveedores, centros de distribución y puntos de venta con el objetivo de reducir costos y tiempos de entrega.
Esto tracciona beneficios y avances en múltiples direcciones, pues permite planificar producción, optimizar la inversión, detectar en tiempo real cambios de hábitos o gustos ajustando decisiones, reduciendo la sobreproducción y stock muerto, contribuyendo a la eficiencia y sostenibilidad.
En la cadena de suministro, optimiza rutas de transporte, evalúa proveedores, predice y anticipa interrupciones de rutas y recomienda hubs logísticos. Con la tecnología RFID (Identificación por radiofrecuencia) permite la visibilidad en tiempo real de los niveles de inventario, rastrear la mercadería, prevenir falsificaciones, realizar control de calidad, y reducir robos y fraudes, pues facilita la trazabilidad y seguridad.
En el eslabón de la comercialización, los clientes reciben recomendaciones de compra según sus hábitos, comportamientos, selección de productos, consultas de talles y pueden obtener una visualización avanzada a través de la realidad aumentada, probador y espejos virtuales o gemelos digitales.
Distintas marcas utilizan la IA para desarrollar soluciones ecológicas, materiales reciclables o procesos de producción que generen menos desechos. Encontramos por ejemplo sistemas inteligentes de gestión energética en tiendas y oficinas que permiten reducir el consumo eléctrico y aplicaciones que calculan la huella de carbono.
Este entramado muestra que la inteligencia artificial no es un accesorio marginal, sino una infraestructura que atraviesa la cadena de valor de la moda.
Todos estos sistemas de inteligencia artificial que predicen (IA predictiva) o generan y crean (IA generativa) están conformados por algoritmos y estos entrenados con grandes volúmenes de datos históricos y patrones de comportamiento, que si bien permiten las optimizaciones enunciadas, también pueden reproducir y amplificar sesgos vinculados con distintos estereotipos y en el caso de la moda estos pueden ser de belleza, discriminación por género, raza, talla, edad u orientación sexual, entre otras dimensiones.
Al mismo tiempo, la creciente presión sobre las empresas para acreditar compromisos con la sostenibilidad y con criterios ESG abre la puerta a nuevas formas de opacidad y de greenwashing algorítmico, en las que la complejidad técnica de los sistemas se utiliza como pantalla para justificar decisiones difícilmente auditables. En este contexto, la pregunta por la justicia y ética algorítmica en la industria de la moda y diseño se vuelve ineludible.

2. Sesgos Algorítmicos en la cadena de valor de la industria de la moda

La IA opera y aprende basada en los datos que recibe.
Está programada a través de algoritmos que "son una serie de instrucciones que deben ser ejecutadas en forma automática por un ordenador".[1] Uno de los más importantes es el llamado "aprendizaje automatizado" o "*machine learning*", que usa instrucciones para encontrar patrones de datos. Puede ser *supervisado* (se le dan ejemplos etiquetados), *no supervisado* (se le exige que busque por sí misma los patrones) y *de refuerzo*, la máquina decide según su prueba o error. Nacido de la combinación de los algoritmos de aprendizaje automático con las redes neuronales formales y con el uso de los macrodatos, encontramos el "deep learning" o "aprendizaje profundo" que revolucionó la inteligencia artificial.[2]
Utiliza redes neuronales artificiales compuestas con muchas capas de nodos (neuronas) que procesan los datos, cada nodo y capa pasa la información a la siguiente. Las redes profundas tienen muchas capas que van pasando y acumulando información y procesándola, y son capaces de realizar tareas complejas, como reconocer imágenes o entender el lenguaje humano.
Los algoritmos entonces, se entrenan con grandes volúmenes de datos. Si los datos de entrenamientos son sesgados, los resultados también lo serán. Esto no es inocuo. En el caso de la moda esto podría significar que ciertos estilos, cuerpos o colores sean favorecidos mientras otros excluidos o estigmatizados. El desafío lo encontramos en cómo evitar que los algoritmos perpetúen estereotipos de belleza talla o raza, los cuales pueden crear una exclusión social, aumento de discriminación, limitar la diversidad y la inclusión de un sector.

2.1 ¿Que es un sesgo?

El sesgo no es exclusivo de la IA. Está presente en la sociedad: desigualdades históricas, sociales e institucionales que luego se "cargan" en los datos.
En el lenguaje común, el sesgo implica tendencia, inclinación hacia un lado. También se lo empareja con el prejuicio y ya más estadísticamente como una diferencia sistemática entre una muestra y la población real.
El diccionario de la Real Academia Española en su acepción 7, lo define como: "Error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras."[3]
Existen distintas clasificaciones de sesgos y subsesgos y más allá de cuál sea la que se adopte, es importante tener en cuenta que se combinan y alimentan entre si a lo largo de todo el ciclo de vida de los sistemas de IA. Dicho ciclo de vida comprende desde la investigación, la concepción y el desarrollo hasta el despliegue y la utilización, pasando por el mantenimiento, funcionamiento, comercialización, financiación, seguimiento, evaluación, validación, y el fin de la utilización, desmontaje y retiro del sistema.

2.2. Categorías y Subcategorías de Sesgos Algorítmicos[4]

Tomando la categorización que realiza El Instituto Nacional de Estándares y Tecnología (NIST por sus siglas en ingles) [5] del Departamento de Comercio de Estados Unidos en su publicación especial 1270 "Hacia un estándar para la identificación y Gestión del sesgo en la inteligencia artificial", encontramos tres categorías troncales con sus subcategorías:

1) ***Sesgo Sistémico***: Histórico, Social e Institucional;

2) ***Sesgo Estadístico Computacional***: 2.*1)* ***En la selección y muestreo de datos***: sesgo de generación de datos; de generalización; de detección; de falacia ecológica; de evaluación; de exclusión; de medición; de popularidad; de población; de representación; de paradoja de Simpson; temporal, de incertidumbre. *2.2.)* ***En el proceso de validación de datos***: sesgo de amplificación, de propagación del error; sesgo heredado; de selección del modelo; de supervivencia; y 2. 3) ***En el uso e interpretación de los sistemas de IA***: desviación conceptual; sesgo de actividad; emergente; de producción de contenido; dragado de datos; bucle de retroalimentación y de vinculación.

3) ***Sesgo Humano***: los cuales se agrupan en: 3.1) ***Individuales***: sesgo de complacencia; consumidor; confusión modal; cognitivo; anclaje; heurística de disponibilidad; confirmación; Dunning-Krueger; implícito; perdida de conciencia de situación; interacción con el usuario; conductual; interpretación; Rashomon; adherencia selectiva; farola; informe anotador; informe humano; presentación y clasificación; 3.2) ***Grupales***: pensamiento grupal; financiación; despliegue y falacia de costo hundido.

Esta estructura resulta particularmente útil para analizar el funcionamiento de la IA en general y en la industria de la moda y el diseño en particular, en la medida en que permite articular la discusión técnica con las dimensiones históricas, sociales y jurídicas de la discriminación algorítmica.

2.2.1 Sesgos sistémicos

Los sesgos sistémicos se originan en estructuras de desigualdad que anteceden a la tecnología. Comprenden patrones de racismo, sexismo, clasismo y otras formas de discriminación que se han consolidado históricamente.

En la moda estos sesgos se manifiestan en cánones de belleza que privilegian cuerpos jóvenes, delgados y blancos, en políticas de talles que excluyen sistemáticamente a personas de peso y en la precarización de determinados segmentos de la cadena de producción. Cuando se integran datos provenientes de estas realidades desiguales en sistemas de inteligencia artificial, los sesgos sistémicos se traducen en decisiones automáticas que refuerzan la exclusión. Por ejemplo: bases de imágenes con poca diversidad corporal pueden conducir a algoritmos de selección de campañas que eligen casi siempre modelos con características hegemónicas.

Subcategoría de sesgos sistémicos:

- ***Sesgo histórico*** se refiere a los prejuicios arraigados en la sociedad a lo largo del tiempo. Relacionado con, pero distinto, los prejuicios en la descripción histórica, o la interpretación, el análisis y la explicación de la historia. Un ejemplo común de sesgo histórico es la tendencia a ver el mundo desde una perspectiva occidental o europea.[5] Los modelos de IA generativa de moda (para marketing o prototipado) se entrenan con vastas bases de imágenes históricas y actuales. Si estas bases de datos están históricamente sesgadas hacia cuerpos jóvenes, delgados y blancos, el algoritmo aprenderá y priorizará estas características, invisibilizando otras etnias, tallas, o edades en imágenes de salida.

• ***Sesgo social*** puede ser positivo o negativo y adoptar diversas formas, pero generalmente se caracteriza por estar a favor o en contra de grupos o individuos según identidades sociales, factores demográficos o características físicas inmutables. Los sesgos sociales suelen ser estereotipos. Ejemplos comunes de sesgos sociales se basan en conceptos como raza, etnia, género, orientación sexual, nivel socioeconómico, educación, etc. El sesgo social se reconoce y analiza a menudo en el contexto de los modelos de PNL (Procesamiento del Lenguaje Natural)[5]. Un ejemplo llevado a la industria de la moda podría ser un modelo influencer scoring que otorga mayor puntaje a mujeres heteronormativas y de clase media/alta y relega a influencers LGBTIQ+, plus size o de barrios populares.
• ***Sesgo institucional***: a diferencia de los sesgos que se manifiestan a nivel de personas individuales, el sesgo institucional se refiere a una tendencia que se manifiesta a nivel de instituciones enteras, donde las prácticas o normas resultan en el favorecimiento o desventaja de ciertos grupos sociales. Ejemplos comunes incluyen el racismo y sexismo institucional.[5]

2.2.2 Sesgos estadísticos y computacionales
Los sesgos estadísticos y computacionales aparecen cuando los datos con los que se entrenan y validan los sistemas de IA no representan adecuadamente la realidad sobre la que se tomarán decisiones. Incluyen la falta de cobertura de ciertos grupos, la exclusión de variables relevantes, errores de medición, usos de métricas inadecuadas y fenómenos como la paradoja de Simpson.

Subcategoría de sesgos estadísticos y computacionales:

2.2.2.1 Sesgos en la selección y muestreo de datos:
• ***de generalización o generación de datos***: se produce por la adición de muestras de datos sintéticos o redundantes a un conjunto de datos[5]. Al aumentar o complementar los datos se genera información sintética o redundante que refuerza un patrón dominante haciendo que el modelo de IA crea que "eso" es la norma para todos los casos, aunque no lo sea.
• ***de detección***: refiere a las diferencias sistemáticas en la forma en que se determinan los resultados entre grupos que pueden causar una sobreestimación o subestimación del tamaño del efecto.[5] Un ejemplo podría constituirlo un sistema que solo recoge reseñas de clientas que compran on-line, ignorando experiencias de locales físicos, distorsionando la percepción de calidad y talles.
• ***de falacia ecológica***: ocurre cuando se hace una inferencia sobre un individuo basándose en su pertenencia a un grupo.[5] Un ejemplo es concluir que las mujeres de un determinado país o región prefieren ropa ajustada porque ese producto vende mucho allí sin ver que solo determinados segmentos acceden a esa oferta.
• ***de evaluación***: surge cuando las poblaciones de prueba o de referencia externa no representan de manera equitativa las distintas partes de la población de usuarios o por el uso de métricas de rendimiento que no son apropiadas para la forma en que se utilizará el modelo[5]. Ej. uso de métricas de precisión promedio en sistemas de recomendación de talles, sin analizar errores concentrados en minorías corporales (ej. Talles grandes).

- ***de exclusión***: Cuando se excluyen grupos específicos de poblaciones de usuarios de las pruebas y análisis posteriores.[5] En estos supuestos la recomendación del sistema o modelo de IA será errática para los grupos excluidos o directamente inexistente.
- ***de medición***: Se presenta cuando las características y las etiquetas son sustitutos de las cantidades deseadas, lo que puede dejar de lado factores importantes o introducir ruido de grupo o dependiente de la entrada que conduce a un rendimiento diferencial.[5] Ej. usar devolución en 7 días como proxy de incomodidad de la prenda y penalizar diseños innovadores que se devuelven por errores de comunicación o color en foto, no por confort.
- ***de popularidad:*** es una forma de sesgo de selección que ocurre cuando los artículos que son más populares están más expuestos y los artículos menos populares están subrepresentados.[5] Esto puede terminar perjudicando a diseñadores emergentes, productos sustentables menos masivos o más costosos o a estéticas minoritarias.
- ***de población:*** Distorsiones sistemáticas en la demografía u otras características de los usuarios entre una población de usuarios representada en un conjunto de datos o en una plataforma y alguna población objetivo[5] Aquí el modelo aprende sobre un tipo de usuarios, pero se usa para decidir sobre otro distinto.
- ***de representación***: Surge debido a un muestreo no aleatorio de subgrupos, lo que provoca que las tendencias estimadas para una población no sean generalizables a los datos recopilados de una nueva población.[5] El dataset se encuentra entonces desbalanceado.
- ***efecto de paradoja de Simpson***: fenómeno estadístico en el que la asociación marginal entre dos variables categóricas es cualitativamente diferente de la asociación parcial entre las mismas dos variables después de controlar una o más variables adicionales. Por ejemplo, la asociación o correlación estadística detectada entre dos variables para una población completa desaparece o se invierte al dividir la población en subgrupos.[5] Ejemplo: a nivel global parece venderse más talles pequeños, al segmentar por canal se ve que online los talles grandes se agotan y hay demanda insatisfecha pero el modelo decide reducir su producción.
- ***temporal***: surge de las diferencias en las poblaciones y los comportamientos a lo largo del tiempo.[5] Esto provoca un modelo desactualizado que arroja respuestas erradas que ocasionara decisiones equivocadas.
- ***de incertidumbre***: cuando los algoritmos predictivos favorecen a los grupos que están mejor representados en los datos de entrenamiento, ya que habrá menos incertidumbre asociada con esas predicciones.[5] Aquí se produce lo que se conoce como una discriminación del modelo por prudencia y el sistema otorgara recomendaciones, respuestas, sugerencias a los grupos con mayores datos mientras que a los otros no se les mostrará por ejemplo productos nuevos, líneas premium, invitaciones y puede desatenderse mercados importantes por falta de datos no por falta de interés.

2.2.2.2 Sesgos en el proceso de validación de datos:

- ***de amplificación*:** este tipo de sesgo se presenta cuando la distribución de los resultados de la predicción está desalineada con respecto a la distribución previa del objetivo de la predicción.[5] Ej. algoritmos que detectan mayor devolución en talles grandes y recomiendan reducir su producción consolidando exclusión de consumidores con cuerpos diversos, lo que produce un riesgo de invisibilización de grupos vulnerables bajo indicadores exitosos.

- ***heredado/sesgo de propagación de error***: surge cuando las aplicaciones desarrolladas con aprendizaje automático se utilizan para generar entradas para otros algoritmos de aprendizaje automático. Si la salida presenta algún sesgo, este puede ser heredado por sistemas que la utilizan como entrada para aprender otros modelos.[5] Esto expande y perpetua el sesgo, y con ello el daño o no beneficio inherente al mismo.
- ***de selección del modelo***: El sesgo introducido al usar los datos para seleccionar un único modelo aparentemente "mejor" de un amplio conjunto de modelos que emplean múltiples variables predictoras. El sesgo de selección de modelo también ocurre cuando una variable explicativa tiene una relación débil con la variable de respuesta.[5] Esto ocurre cuando para por ejemplo una predicción determinada, usando lo mismos datos o combinándolos, se prueba con distintos modelos o sistemas de IA y se termina seleccionando al modelo que se ajusta demasiado a esas variables introducidas y no tanto al fenómeno general.
- ***de supervivencia***: refiere a la tendencia de las personas a centrarse en los elementos, observaciones o personas que "sobreviven" o superan un proceso de selección, mientras que pasan por alto las que no lo hicieron.[5] Ej. analizar que diseñadores independientes tienen éxito en un Marketplace sin considerar a quienes nunca pudieron entrar por requisitos técnicos o económicos y entrenar la IA solo con ese subconjunto.

2.2.2.3 Sesgos en el uso e interpretación de los sistemas de IA:

- ***sesgo de actividad*:** este tipo de sesgo de selección se manifiesta cuando los sistemas o plataformas entrenan sus modelos utilizando datos provenientes mayormente de sus usuarios más activos, dejando de lado a aquellos que son menos activos o están inactivos.[5] Ej. personalización que funciona perfecto para usuarias que interactúan a diario con la app, pero casi no atiende a personas que compran esporádicamente o con poca alfabetización digital.
- ***sesgo desviación conceptual/emergente***: Uso de un sistema fuera del dominio de aplicación planificado y una causa común de brechas de rendimiento entre los entornos de laboratorio y el mundo real.[5] Aquí el problema no se encuentra tanto en los datos de uso sino en el contexto de uso, haciéndolo en uno distinto para el que fue diseñado sin ajustarlo o revalidarlo.
- ***de producción de contenido***: se origina debido a las diferencias estructurales, léxicas, semánticas y sintácticas que se encuentran en los contenidos generados por los usuarios.[5] Ej. recomendador entrenado con fotos de looks subidas sobre todo por influencers urbanos de alto poder adquisitivo, la moda de barrios populares casi no aparece en las recomendaciones.
- ***dragado de datos***: un sesgo estadístico en el que probar un gran número de hipótesis de un conjunto de datos puede parecer producir significancia estadística incluso cuando los resultados no son estadísticamente significativos[5]. Aquí se produce un ruido disfrazado de un falso patrón. Se confunde un efecto casual encontrado después de buscar e intentar masivamente en el dataset con una verdad robusta sobre las preferencias de los consumidores que difícilmente responda a una hipótesis razonable previa.
- ***bucle de retroalimentación***: ocurre cuando un algoritmo aprende del comportamiento del usuario y retroalimenta ese comportamiento en el modelo.[5] Ej. el algoritmo muestra más talles estándar, se venden más esos talles, entonces el sistema concluye que "son los más deseados".

• ***de vinculación***: Surge cuando los atributos de red obtenidos de las conexiones, actividades o interacciones de los usuarios difieren y tergiversan su verdadero comportamiento. Ej. sistema que ubica en la misma "comunidad de estilo" a quienes siguen las mismas cuentas de lujo, sin distinguir diferencias culturales o económicas y recomienda productos inaccesibles a parte del grupo.

2.2.3 Sesgos humanos en el diseño y uso de sistemas

Por su parte, los sesgos humanos incluyen atajos cognitivos y errores de juicio de quienes diseñan, entrenan y usan sistemas de inteligencia artificial. Entre ellos el sesgo de confirmación, la tendencia a confiar de manera excesiva en la decisión automatizada, el efecto Dunning Kruger y lo que algunos autores denominan efecto Rashomon, donde modelos distintos ofrecen explicaciones diferentes para un mismo fenómeno y se elige aquella que mejor se ajusta a intereses o intuiciones previas.

En la moda estos sesgos pueden conducir a decisiones que se atribuyen al algoritmo pero que en realidad responden a preferencias de direcciones creativas o de equipos de marketing. Por ejemplo, cuando los resultados de un sistema de análisis de tendencias se interpretan de manera selectiva para justificar la repetición de una estética excluyente, o cuando se descartan advertencias sobre sesgos en los datos por considerarlas incompatibles con la imagen de marca.

Subcategorías de Sesgos humanos en el diseño y uso de sistemas

2.2.3.1 Individuales:

• ***de complacencia a la automatización***: se produce cuando los seres humanos confían demasiado en los sistemas automatizados o cuando sus propias habilidades se ven atenuadas por dicha dependencia excesiva. Las personas asumen que los sistemas automatizados son más precisos, consistentes y confiables que los humanos y difieren las decisiones y sugerencias hacia ellos. Un ejemplo de este sesgo es la sobre confianza en la corrección automática de correctores ortográficos y gramaticales.[5] Ej. equipos de marketing confían ciegamente en sistemas de predicción de tendencias, ignorando advertencias sobre sesgos en los datos.

• ***del consumidor***: se produce cuando un sistema o aplicación proporciona a los usuarios un nuevo lugar para expresar sus sesgos, y puede ocurrir desde cualquier lado, o parte, en una interacción digital.[5] En este caso los sesgos del usuario se cargan al sistema que abre un canal a traves de reseñas, reportes, likes, valoraciones, etc y luego se incorporan y el sistema lo toma como señal de "lo que funciona" o "lo que es deseable".

• ***de confusión modal***: cuando las interfaces modales confunden a los operadores humanos, quienes no comprenden qué modo utiliza el sistema y toman acciones correctas para un modo diferente, pero incorrectas para su situación actual. Esto es la causa de muchos accidentes mortales, pero también una fuente de confusión en la vida cotidiana.[5] Ej. panel de diseño asistido por IA que mezcla sin aclaración modos "borrador" y "aprobado"; los creativos interpretan como definitivas propuestas que aun eran experimentales.

• ***Cognitivos:*** término amplio que se refiere generalmente a un patrón sistemático de desviación del juicio racional y la toma de decisiones. A lo largo de muchas décadas de investigación sobre el juicio y la toma de decisiones, se ha identificado una gran varie-

dad de sesgos cognitivos, algunos de los cuales son atajos mentales adaptativos conocidos como heurísticas.[5] Ej. equipo creativo que interpreta salidas de la IA según sus prejuicios estéticos reforzando la idea de que la moda que vende es la que responde a un único tipo de cuerpo y estilo.

- ***de anclaje***: es una especie de sesgo cognitivo relacionado con la influencia de un determinado punto de referencia o anclaje en las decisiones de las personas. Una vez que se coloca un ancla, las personas se ajustan desde ese punto de anclaje para llegar a una respuesta final. Los tomadores de decisiones suelen estar sesgados hacia un valor presentado inicialmente.[5] Ej. una plataforma de venta en línea usa un algoritmo de precios dinámicos para optimizar el margen, si el algoritmo de fijación de precios esta inicialmente anclado a un precio de referencia alto basado en el de un competidor premium no relevante, incluso cuando las variables del mercado (demanda, costos) sugieran una reducción significativa, el sistema de IA tenderá a realizar solo pequeños ajustes desde ese ancla inicial. Esto puede resultar en precios incorrectos o en la pérdida de competitividad de la marca, debido a la influencia irracional del dato inicial.
- ***heurística de disponibilidad:*** se trata de un atajo mental por el cual las personas tienden a dar mayor relevancia a aquello que le viene fácil o rápidamente a la mente. Por ello, lo que es más fácil de recordar —es decir, está más disponible— recibe un mayor énfasis en el juicio y la toma de decisiones.[5] Las consecuencias de este sesgo es que se sobredimensionan episodios llamativos (un escándalo, una campaña fallida, una experiencia personal fuerte) con la correlativa minimización de patrones estadísticos que son menos llamativos, pero más frecuentes.
- ***de confirmación***: es una especie de sesgo cognitivo en el que las personas tienden a preferir información que se alinee con, o confirme, sus creencias existentes. Las personas pueden exhibir un sesgo de confirmación en la búsqueda, interpretación y recuerdo de información.[5] Ej. directores creativos interpretan resultados de IA para justificar estéticas excluyentes ya definidas, esto trae aparejado el riesgo de usar la IA como argumento de autoridad para legitimar decisiones discriminatorias.
- ***efecto dunning-kruger***: representa la tendencia de las personas con baja capacidad en un área o tarea determinada a sobreestimar su propia capacidad. Generalmente se mide comparando la autoevaluación con el desempeño objetivo, a menudo denominados capacidad subjetiva y capacidad objetiva, respectivamente[5]**.** Ej. diseñadores con escaso conocimiento técnico en IA sobreestiman su capacidades para interpretar resultados, aplicando decisiones erróneas. Dirección creativa sin formación en datos que cree entender la IA y usa modelos sin consultar a especialistas, lanzando colecciones altamente sesgadas bajo la etiqueta de "diseño inteligente".
- ***implícito:*** creencia, actitud, sentimiento, asociación o estereotipo inconsciente que puede afectar la forma en que los humanos procesan la información, toman decisiones y realizan acciones.[5] Ej. anotadores humanos que etiqueta imágenes de moda tienden a clasificar prendas según estereotipos de género o raza.
- ***perdida de conciencia situacional***: cuando la automatización lleva a que los humanos no sean conscientes de su situación, de modo que, cuando se les devuelve el control de un sistema en una situación en la que los humanos y las máquinas cooperan, no están preparados para asumir sus funciones. Esto puede ser una pérdida de conciencia sobre lo que

la automatización está y no está solucionando.[5] Ej. la marca delega el control de precios y stock en sistemas automáticos sin detectar el impacto social de reducir talles o retirar líneas inclusivas provocando un cuestionamiento a su reputación.

- ***de interacción con el usuario***: se produce cuando un usuario impone sus propios sesgos y comportamientos autoseleccionados durante la interacción con los datos, la producción, los resultados, etcétera[5]**.** Ej. creativos que solo exploran las primeras opciones que arroja el generador de diseños y no revisan alternativas menos visibles limitando la diversidad estética. En este caso el propio diseño del sistema puede contener los sesgos, por ejemplo si dentro de las 20 opciones que se le presentan al usuario las 10 primeras son prendas de una estética concreta y para ver otras estéticas tiene que descender el usuario a las ultimas opciones, puede ocurrir que no lo haga, que no sea su modo de interactuar, que no quiera invertir tanto tiempo, etc. El sistema usara los primeros clics como señal de preferencia.
- ***de interpretación***: es un sesgo en el procesamiento de la información que puede ocurrir cuando el usuario interpreta el resultado algorítmico de acuerdo a sus propios sesgos y visiones.[5] Se proyecta el prejuicio propio sobre un resultado ambiguo y se usa el informe algorítmico como argumento para justificarlo.
- ***efecto de rashomon***: se refiere a las diferencias en la perspectiva, memoria y recuerdo, interpretación e informes sobre el mismo evento de múltiples personas o testigos.[5] Ejemplo: las diferentes interpretaciones dentro de una misma empresa de los distintos equipos: (i) equipo de marketing interpreta los resultados como una señal para lanzar una línea juvenil con estética de uso en las calles enfocada en mercados urbanos europeos; (ii) equipo de diseño considera que la tendencia refleja una búsqueda de comodidad y propone una colección sin género (que puede ser usada por cualquier persona) con tejidos suaves y cortes amplios; (iii) por su parte el equipo de Dirección creativa cree que el algoritmo está captando una moda pasajera y decide mantener la línea clásica de la marca y (iv) Consultores externos en diversidad advierten que el sistema está sesgado por la sobrerepresentación de influencers de determinada región, y que la "tendencia" podría invisibilizar otros estilos. El resultado arrojará que aunque todos analizan el mismo output del sistema, sus conclusiones son radicalmente distintas. El sistema se convierte en un espejo de sus propias creencias, intereses o intuiciones. Esto es el efecto Rashomon: múltiples verdades construidas sobre el mismo evento algorítmico.
- ***de adherencia selectiva***: es la inclinación de los responsables de toma de decisiones a adoptar selectivamente el asesoramiento algorítmico cuando coincide con sus creencias y estereotipos preexistentes.[5] Ej. validación de colecciones de diseño: El director demuestra adherencia selectiva al destacar y actuar solo sobre las predicciones que confirman su visión preexistente y descarta activamente las alertas que sugieren desviarse del estilo ya concebido, obstaculizando la innovación real y gestión de riesgos.
- ***de alumbrado público o semáforo***: un sesgo por el cual las personas tienden a buscar solo donde es más fácil mirar.[5] Ej. analizar solo datos de redes sociales de alto tráfico e ignorar ferias, mercados y canales informales donde se gesta parte importante de la moda local.
- ***de informe del anotador***: cuando los usuarios confían en la automatización como un sustituto heurístico para su propia búsqueda y procesamiento de información. Una forma de sesgo individual, pero a menudo se describe como sesgo de grupo o sus efectos más amplios en los modelos de procesamiento del lenguaje natural.[5]

- ***sesgo de informe humano***: Cuando los usuarios confían en la automatización como un reemplazo heurístico para su propia búsqueda y procesamiento de información.[5] Ej. los anotadores ajustan sus juicios para alinearse con lo que el sistema predice, por temor a parecer "equivocados".
- ***de clasificación***: es una forma de sesgo de anclaje referido a la idea de que los resultados mejor clasificados son los más relevantes e importantes y darán lugar a más clics que otros resultados.[5] Ej. El equipo se fija solo en las tendencias virales identificadas por la IA y descuida micro tendencias de nicho estratégicas para públicos específicos.
- ***Conductual:*** son las distorsiones sistemáticas en el comportamiento de los usuarios en plataformas o contextos, o entre usuarios representados en diferentes conjuntos de datos. Las decisiones se basan en la conducta de un tipo de usuario y contexto, y se trasladan como si fueran naturales al resto, perjudicando a quienes compran de otro modo o interactúan menos.
- ***de presentación***: se producen como consecuencia de cómo es presentada la información en una página web, a través de una interfaz de usuario, debido a la calificación o clasificación de los resultados, o a través de la interacción sesgada y autoseleccionada de los propios usuarios.[5] Ej. Un sitio de e-commerce de moda utiliza un sistema de recomendación algorítmica que ordena los productos según popularidad y calificación de usuarios. Los productos que aparecen en las primeras posiciones reciben más clics y compras, independientemente de su calidad real y los usuarios tienden a interactuar más con esas opciones. Los algoritmos interpretan que esas prendas son "más deseadas" y las siguen mostrando en posiciones privilegiadas..

2.2.3.2 Grupales:

- ***de pensamiento de grupal***: fenómeno psicológico que ocurre cuando las personas de un grupo tienden a tomar decisiones poco óptimas basándose en su deseo de conformarse con el grupo o en el miedo a discrepar. En el pensamiento grupal, las personas a menudo se abstienen de expresar su desacuerdo personal con el grupo, dudando en expresar opiniones que no se alinean con él.[5] Ej. Una marca internacional de moda utiliza un sistema de IA para analizar tendencias globales y recomendar estilos para la próxima temporada. El algoritmo sugiere que los colores neutros y minimalistas son los más populares en redes sociales. El equipo creativo se reúne para interpretar los resultados. Aunque algunos diseñadores creen que la IA está sobrevalorando una tendencia pasajera y que sería más inclusivo apostar por paletas vibrantes inspiradas en otras culturas no se animan a contradecir la visión dominante. El resultado es una colección homogénea, que refuerza estéticas hegemónicas y deja fuera expresiones culturales diversas.
- ***de financiación***: Surge cuando se informan resultados sesgados para apoyar o satisfacer a la agencia de financiación o al patrocinador financiero del estudio de investigación pero también puede ser el investigador individual.[5] Este sesgo (*Funding Bias*), representa una preocupación significativa, ya que la dependencia de fondos privados para la investigación y desarrollo de IA puede comprometer la objetividad científica o técnica de los sistemas. Ej. El proyecto es financiado en su totalidad por un gran conglomerado de *Fast Fashion*, cuya reputación está siendo afectada por críticas sobre el *greenwashing* y la sobreproducción masiva. El equipo de investigación, consciente de su dependencia financiera,

prioriza inconsciente o explícitamente el desarrollo de un módulo de IA que se centre y optimice únicamente en métricas que el patrocinador ya cumple o que puede alcanzar fácilmente.

- ***Implementación***: Surge cuando los sistemas se utilizan como asistentes para la toma de decisiones humanas, dado que el intermediario humano puede actuar sobre las predicciones de formas que normalmente no se modelan en el sistema. Sin embargo, siguen siendo las personas las que utilizan el sistema desplegado.[5] Ej. Una marca de moda utiliza un sistema de IA para micro-segmentar su base de clientes con el fin de optimizar el *marketing* y la promoción de nuevas colecciones basándose en el historial de navegación, compras, geolocalización y datos demográficos disponibles. El modelo asigna puntuaciones de riesgo (baja, media, alta) a cada cliente para una campaña específica, el sistema los clasifica en grupo A clientes de ingresos altos de una zona urbana prestigiosa y grupo B a clientes de ingresos altos de una zona periférica. El Gerente de *Marketing* aplica su propio sesgo implícito optando por destinar los recursos de marketing para el grupo A para asociar la marca a las zonas urbanas prestigiosas.
- ***falacia al costo hundido***. Tendencia humana por la cual las personas optan por continuar con un esfuerzo o comportamiento debido a recursos previamente gastados o invertidos, como dinero, tiempo y esfuerzo, sin importar si los costos superan los beneficios.[5] Ej. podría llevar a los equipos de desarrollo y a las organizaciones a creer que, dado que ya han invertido tanto tiempo y dinero en una aplicación de IA en particular, deben llevarla al mercado en lugar de decidir abandonar el esfuerzo, incluso ante una deuda técnica o ética significativa.
- Los sesgos algorítmicos no son meros defectos técnicos, sino articulaciones complejas entre estructuras de desigualdad, decisiones de diseño estadístico y comportamientos humanos. Su mitigación requiere de un enfoque integral que cubra todo el ciclo de vida del sistema de Inteligencia Artificial (IA). Así lo ha recomendado la UNESCO a través de sus directrices éticas y estos principios se vienen utilizando en la práctica y uso de sistemas de IA e incluso en regulaciones como la de la UE.

2.2.4 Mapa de sesgos a lo largo de la cadena de valor

La combinación de sesgos sistémicos, estadísticos y humanos se despliega a lo largo de toda la cadena de valor de la moda asistida por inteligencia artificial.

En la investigación de tendencias se corre el riesgo de amplificar voces dominantes y silenciar expresiones locales.

En el diseño asistido por algoritmos pueden reproducirse estéticas excluyentes.

En la producción y en la logística se optimizan costos sin considerar impactos laborales.

En el marketing y en el comercio electrónico los sistemas de recomendación pueden segmentar a consumidores de manera discriminatoria.

Finalmente, en la etapa de sostenibilidad, los paneles de datos pueden ocultar impactos ambientales y sociales relevantes y ello nos lleva a la consideración de lo que se conoce como greenwashing algorítmico.

3. Greenwashing Algorítmico en la industria de la moda

El greenwashing tradicional es el que describe estrategias comunicacionales mediante las cuales una empresa se presenta como ambientalmente responsable sin que esa imagen se corresponda con sus prácticas reales. Con la incorporación de sistemas de inteligencia artificial emerge una variante específica que puede denominarse greenwashing algorítmico. Se produce cuando paneles de datos, etiquetas automatizadas, puntuaciones y visualizaciones sintetizan de manera engañosa la información ambiental o social y generan una apariencia de sostenibilidad basada en datos incompletos o poco verificables.
En la moda esto puede observarse cuando aplicaciones asignan sellos verdes a productos sin detallar criterios, cuando se muestran reducciones porcentuales de impacto calculadas sobre líneas de base poco claras o cuando se dirigen mensajes de sostenibilidad solo a determinados segmentos de consumidores. La combinación de opacidad técnica y narrativa de datos objetivos dificulta el control por parte de organismos de supervisión y de la sociedad civil.
Representa un blanqueo verde digital apoyado en tecnología, automatización y datos manipulados que genera percepción falsa, sesgada o idealizada de sostenibilidad. Son algoritmos que presentan productos como sostenibles aunque no lo sean; etiquetado automático o predictivo como ecofriendle por IA, datos manipulados, sesgados o incompletos en dashboards, marketplaces, IA.
Para neutralizar o disminuir el greenwhasing algorítmico, como con los sesgos en general, no alcanza con prohibir publicidad engañosa, se requiere intervención integral técnica, jurídica, ética, documental y de normativa porque el engaño se genera en la capa de datos, algoritmos, visualización y trazabilidad digital.

4. Hacia una justicia Algorítmica

La justicia algorítmica busca identificar, prevenir y corregir sesgos en sistemas automatizados que afectan derechos o refuerzan desigualdades.
Esta tarea se ha entendido sociotécnica, ya que no puede abordarse exclusivamente desde el error técnico o estadístico, demanda tener en cuenta el contexto social, cultural, económico, entre otros. "*El sesgo no es simplemente un tipo de error técnico, se abre a las creencias humanas, los estereotipos, las formas de discriminación*"[6]

5. Mitigación de riesgos de sesgos algorítmicos. Respuestas Técnicas, Recomendaciones, Resoluciones.

Joy Buolamwini, fundadora de Algorithmic Justice League, activista afroamericana contra sesgos en IA, se refirió a la necesidad de la "*responsabilidad en la codificación*". Trabajaba en un software de análisis facial cuando descubrió que no detectaba su rostro porque quienes codificaron el algoritmo no le habían enseñado a identificar una alta gama de tonos de piel y

estructuras faciales. Se propuso entonces combatir el sesgo en el aprendizaje automático, un fenómeno que llamó "*mirada codificada*" que puede propagar sesgos a gran escala, experiencias de exclusión, y prácticas discriminatorias. Para ello abordó posibles caminos: pensar en un código más inclusivo y empleo de prácticas de codificación inclusivas, auditar, crear grupos de formación integradores y pensar en el impacto de las tecnológicas que se están creando. Propuso prestar atención a ¿quiénes codifican? ¿Cómo? y ¿por qué?[7]
Para mitigar los sesgos se están tomando medidas técnicas, éticas, regulatorias y organizativas sobre el consenso de que el hombre, junto al resto de los seres vivos, medioambiente y ecosistema, resulta el centro de protección.

5.1 La Aplicación de Principios y Valores sobre la ética algorítmica
La **UNESCO** adoptó en 2021 la "*Recomendación sobre la Ética de la Inteligencia Artificial*". Ya había hecho lo propio en 2020 el Consejo de la **Unión Europea (UE)** a través del "*Libro Blanco sobre IA*"[8] y en 2019 "*Directrices éticas para una inteligencia artificial fiable*"[9] *como asimismo la* ***OCDE*** *(Organización para la Cooperación y el Desarrollo Económicos) en 2019 revisada y actualizada en 2023/2024* en *la "Recomendación del Consejo sobre Inteligencia Artificial*[10] .
La Recomendación de la UNESCO establece principios para garantizar que la IA beneficie a toda la humanidad, recomendándolos en todo el ciclo de vida de los sistemas o modelos. Se apoya en el respeto, la protección y promoción de los derechos humanos, las libertades fundamentales, la dignidad humana, la inclusión y la sostenibilidad. A partir de este marco se han llevado a cabo periódicamente estudios y seguimiento sobre la preparación de los distintos países para aplicar estas recomendaciones a través del "Readiness Assessment Methodology"(RAM) que constituye una herramienta para ayudar a los Estados a identificar brechas regulatorias, institucionales y técnicas, y orientar la construcción de ecosistemas de IA éticos y responsables.[11,12]
Por su parte, en el documento normativo "Declaración de Bletchley" efectuada en el marco de la Cumbre de Seguridad de la IA, en noviembre de 2023, actualizado en febrero 2025[13] se hizo hincapié en la necesidad de la transparencia, explicabilidad, supervisión humana apropiada, rendición de cuentas, enfoque ético en el uso de los sistemas, en la promoción de la inclusión, la equidad global y la reducción de las brechas tecnológicas para que los beneficios de la IA lleguen a todos y no solo a unos pocos.
Los principios, valores y directrices de estos instrumentos internacionales han sido nutrientes y base de numerosas otras recomendaciones, resoluciones, comités e incluso normativa.[14]
La ONU en septiembre de 2024 en su informe final "*Gobernanza de la IA para la humanidad*" propone establecer una arquitectura global inclusiva y cooperativa para la gobernanza basada en los derechos humanos, el desarrollo sostenible y la paz. Pretende cerrar brechas de representación, coordinación y capacidad, asegurando que la IA beneficie a toda la humanidad y que sus riesgos sean identificados y mitigados. [15]
El denominador común en los distintos instrumentos referidos, es sin duda, la dignidad humana que implica el reconocimiento del valor intrínseco e igual de cada ser humano con independencia de su raza, color, ascendencia, genero, edad, idioma, religión, opiniones políticas, origen nacional, étnico o social, condición económica o de nacimiento, discapacidad o cualquier otro motivo. Ningún ser humano ni comunidad humana debería sufrir daños o sometimientos ya sean de carácter físico, económico, social, político, cultu-

ral o mental durante ninguna etapa del ciclo de vida de los sistemas de IA, debiendo estos mejorar la calidad de vida de los seres humanos. Las personas nunca deberían ser cosificadas, ni su dignidad menoscabada, ni sus derechos humanos y libertades fundamentales ser objeto de violación o abusos.[16]
En todo el ciclo de vida de los sistemas de IA debe garantizarse *la diversidad e inclusión* promoviendo la participación activa de todas las personas o grupos, como la diversidad de las elecciones de estilo de vida, creencias, opiniones, expresiones o experiencias personales incluida la utilización opcional de la IA.

Entre los principios que se enarbolan estos documentos internacionales encontramos:

- ***Proporcionalidad e Inocuidad*** donde la decisión final debería ser adoptada por un ser humano.
- ***Seguridad y Protección*** a través del desarrollo de marcos de acceso a los datos que sean sostenibles, respeten la privacidad y fomenten un mejor entrenamiento y validación de los modelos de IA y que utilicen datos de calidad.
- ***Equidad y no Discriminación*** evitando reforzar o perpetuar aplicaciones y resultados discriminatorios o sesgados a lo largo del ciclo de vida de los sistemas de IA a fin de garantizar equidad, con recursos efectivos contra la discriminación y la determinación algorítmica sesgada, abordar las brechas digitales y de conocimientos.
- ***Derecho a la Intimidad y Protección de Datos.*** La privacidad constituye un derecho esencial para la protección de la dignidad, la autonomía y la capacidad de actuar de los seres humanos debiendo ser respetada, protegida y promovida estableciéndose marcos de protección de datos y mecanismos de gobernanza adecuados, protegidos por los sistemas judiciales. Los algoritmos requieren de evaluaciones del impacto en la privacidad.
- ***Supervisión y Decisión Humana.*** Debe velarse porque siempre sea posible atribuir responsabilidad ética y jurídica a personas físicas o a entidades jurídicas existentes.
- ***Transparencia y Explicabilidad.*** Las personas deberían estar informadas cuando una decisión se basa en algoritmos de IA o se toma a partir de ellos, sobre todo cuando afecta su seguridad o derechos humanos, pudiendo conocer los motivos y tener la posibilidad de presentar sus alegaciones a un miembro del personal de la empresa del sector privado, institución o sector público habilitado para revisar y enmendar la decisión. Este principio proclama que los actores de la IA deberían comprometerse a velar por que los algoritmos desarrollados sean explicables.
- ***Responsabilidad y Rendición de Cuentas.*** Debería garantizarse la auditabilidad y trazabilidad del funcionamiento de los sistemas de IA para intentar solucionar cualquier conflicto con las normas relativas a los derechos humanos y las amenazas al bienestar de medio ambiente y los ecosistemas.
- ***Sensibilidad y Educación*** procurando la alfabetización mediática e informacional y la capacitación.[17]

El conjunto de los instrumentos referenciados ofrecen una mirada ética y de gobernanza global, basada en derechos humanos, equidad y justicia social, pero también con enfoque económico, técnico y operacional.

5.2 Estrategias técnicas

La IA funciona en ocasiones en como una caja negra lo que implica que las decisiones tomadas por los algoritmos no son fácilmente comprensibles o auditables para quien las recibe, por lo cual se los denomina también "sistemas opacos". Esto crea un problema de falta de transparencia, de falta de explicabilidad, especialmente cuando los consumidores son afectados por las recomendaciones, precios dinámicos o incluso la selección de productos, o cuando son excluidos. En Europa, el Reglamento General de Protección de Datos (GDPR) reconoce el derecho a una explicación significativa sobre decisiones automatizadas (art 22) y en el nuevo reglamento de la IA de la UE se exige mayor transparencia para sistemas de alto riesgo.

El opuesto a caja negra o sistemas opacos sería lo que se conoce como sistemas transparentes o explicables, en los que sí es posible comprender el resultado.

Se están desplegando estrategias hacia una IA explicable usando algoritmos simples como árboles de decisión o regresiones cuando sea posible; herramientas de explicabilidad que permiten ver qué variables influyeron en la decisión de un modelo complejo; registro del entrenamiento del modelo; auditorías algorítmicas a través de revisiones independientes de cómo funciona el modelo e interfaces explicativas que den al usuario razones comprensibles, entre otras.

5.3 Estrategias Contractuales y de cumplimiento

Los contratos con proveedores tecnológicos pueden incorporar cláusulas sobre calidad y licitud de los datos, gestión de sesgos, acceso a documentación técnica, límites de uso de los modelos y asignación de responsabilidades por daños. A nivel interno las empresas pueden adoptar políticas de inteligencia artificial responsable, establecer procedimientos de evaluación de impacto y articular estos instrumentos con programas de cumplimiento en materia de consumidores, derechos humanos y medio ambiente.

5.4 Estrategia de Cultura organizacional y formación interdisciplinaria

La efectividad de estas herramientas depende de la cultura organizacional. Es necesario que diseñadores, responsables de datos, equipos de marketing y áreas jurídicas compartan un lenguaje común sobre justicia algorítmica y sobre riesgos éticos de la inteligencia artificial. La formación continua y el diálogo interdisciplinario contribuyen a que las decisiones tecnológicas no se tomen de espaldas a sus implicancias sociales.

5.5 Estrategia de Diversidad, Inclusión y Conciencia

Formar equipos de desarrollo inclusivos lo cual tiende a asegurar la diversidad de género, etnia y disciplinas en los equipos de IA contrarresta sesgos de confirmación e implícito, al introducir perspectivas variadas en la definición del problema y la selección de datos. Asimismo, desarrollar métricas de equidad (*Fairness Metrics*) junto a las métricas de rendimiento.

5.6 Estrategia de Auditoría y Trazabilidad.

Mediante mapeos de la representatividad de los datos (imágenes, textos, historiales de compra, distribución equitativa de tonos de piel, morfologías corporales y contextos culturales) utilizados para entrenar modelos de IA.

5.7 Estrategias para mitigar el greenwhashing algorítmico.
Algunas herramientas son la verificación ambiental de datos a través de certificación de huella, emisiones, materiales, usando las distintas normas de certificación; auditorias de datos externos e independientes con validación por organismos de control, auditores técnicos, cámaras de moda sostenible; la aplicación de *FAIR data principles* (datos encontrables, accesibles, interoperables y reusables con trazabilidad documental)[18]; aplicando los principios de IA ética como el de explicabilidad debiendo el algoritmo justificar porqué etiquetó un producto como eco o sostenible; recurrir el principio *human oversight (*supervisión humana) cuando se aplican etiquetas verdes automatizadas o rankings ESG; asimismo al *bias mitigation* (mitigación de sesgos) depurando sesgos que favorecen productos "eco" sin base real; aplicando el principio traceability (trazabilidad) pudiendo rastrearse cada decisión algorítmica hasta el dato ambiental de origen y el principio accountability (responsabilidad) identificando responsables: desarrollador, proveedor de datos, o comercializador; evitar el green storytelling digital sustituyendo promesas por datos verificables; entre otras.
Una perspectiva crítica sobre métricas ESG automatizadas demanda preguntarse quién define los indicadores, qué datos se consideran aceptables y qué márgenes de error se tolera para cada dimensión evaluada.

5.8 Estrategia de gobernanza o regulación
Existen argumentos a favor y en contra de la regulación de la IA, esto sin perjuicio del uso y aplicación de las directrices y recomendaciones sobre el uso ético. Entre los primeros encontramos la intención de proteger los derechos humanos fundamentales y evitar discriminaciones, vigilancia masiva y manipulaciones; asimismo en la intención de asegurar transparencia y explicabilidad, de tender a consolidar la seguridad y confiabilidad; marcar pautas sobre responsabilidad legal identificando responsables ante fallos o sesgos; alcanzar la igualdad de condiciones y evitar o disminuir la brecha digital.
Entre los argumentos en contra está el riesgo al freno a la innovación por exceso de burocracia; la rigidez ante tecnologías dinámicas, pudiendo tornarse obsoletas las regulaciones antes de aplicarse; intervención estatal indebida, sobre control, hipervigilancia; fragmentación global: fronteras algorítmicas o zonas de regulación desigual.
La regulación de la Unión Europea[19], como asimismo proyectos parlamentarios de varios países, han encarado una regulación o proponen hacerlo desde la perspectiva del riesgo. Desde este parámetro, se tiende a regular la IA que genere alto riesgo, prohibir la de inaceptable riesgo y efectuar recomendaciones y directrices sobre la de bajo y limitado riesgo.
Puede ensayarse la siguiente clasificación de la IA desde dicha perspectiva:

- **Riesgo bajo:** sus recomendaciones, usos, respuestas, no representan riesgos significativos.
- **Riesgo limitado:** pueden influir en decisiones humanas y necesitan transparencia (explicabilidad).
- **Riesgo alto:** pueden afectar significativamente la vida de las personas (sobre todo sector sanitario, financiero y justicia).
- **Riesgo inaceptable:** constituye una amenaza para la seguridad, para los derechos fundamentales del hombre o valores democráticos;

Algunos ejemplos en la industria de la moda:

- ***Riesgo Bajo:*** recomendación de outfits, o combinación de ropa que surgieren looks basados en gustos, clima, tendencias; chatbots simples; herramientas generadoras de diseños o moodboards que crean inspiración visual o bocetos que sirven de apoyo creativo pero el diseñador es quien toma la decisión final; análisis de tendencias basado en redes, tiene valor comercial pero bajo riesgo ético y legal; clasificación de productos o etiquetas automáticas en e-commerce, por color, tipo, estilo, temporada.
- ***Riesgo Limitado:*** la utilización de cámaras más IA que revisan costuras, estampados, materiales, para detección de falsificaciones o control de calidad que pueden impactar en decisiones de producción. Herramientas que predicen demandas o comportamiento de compras pueden generar pérdidas económicas o exceso de inventario en caso que fallen; La IA que segmenta a usuarios, clasificando por género, poder adquisitivo, hábitos de compra, pueden generar sesgos o exclusión si no hay una responsable y diligente gestión y transparencia.
- ***Riesgo Alto:*** Los sistemas de reconocimiento facial en tiendas físicas para seguimiento de clientes (riesgo alto por temas de privacidad, vigilancia masiva o discriminación). IA que selecciona candidatos automáticamente para trabajos en moda (modelaje, diseño, ventas) si hay sesgos de género, edad, raza, puede discriminar si no hay supervisión humana. La IA generativa que crea modelos virtuales con estéticas corporales irreales, si no se advierte que han sido generados con IA puede impactar en la salud mental de consumidores, perpetuar estereotipos hegemónicos de apariencia, tallas, etcétera.
- ***Riesgo Inaceptable:*** Sistemas de puntuación social de clientes o empleados lo cual sería discriminatorio y éticamente inaceptable; uso encubierto de reconocimiento facial para perfilamiento (cámaras en tiendas que reconozcan identidades sin consentimiento explicito, cruzando datos con redes sociales o antecedentes de compras) afectando la privacidad y generando discriminación.

En otras regiones, a diciembre de 2025, la tendencia combina marcos éticos con las primeras leyes específicas de IA. China adoptó regulaciones sobre "deepfakes" y servicios de IA generativa, y en 2025 aprobó estándares nacionales de IA, configurando un modelo fuertemente estatal orientado al control de contenidos y a la seguridad pública.[20] Japón mantiene un enfoque de "soft law", sin embargo, en mayo de 2025 sancionó su primera ley marco, la *AI Promotion Act*, que fija principios para el desarrollo responsable de la IA y se complementa con directrices no vinculantes[21]. Corea del Sur, por su parte, aprobó la *AI Basic/Framework Act* (2024–2025), que entrará en vigor en 2026 y establece un marco nacional que busca equilibrar competitividad e intereses de los ciudadanos. En Canadá el proyecto de *Artificial Intelligence and Data Act (AIDA)*, continúa en trámite, mientras que el gobierno ha adoptado un Código de Conducta voluntario para sistemas generativos avanzados como medida transitoria. En Estados Unidos se combinan normas sectoriales (como el *California Consumer Privacy Act/CPRA*) con documentos programáticos de autorregulación aunque la Orden Ejecutiva dictada bajo la anterior administración ha quedado derogada por la actual ante el proyecto de cambio de políticas de desregulación y de remoción de barreras frente a la IA. En América Latina se discuten estrategias nacionales y

proyectos normativos y varios países de la región —entre ellos Ecuador, Chile, Brasil, México o Uruguay— participan en la implementación de la Recomendación de la UNESCO sobre la ética de la IA y en el ejercicio de evaluación RAM; Ecuador ya ha completado su informe RAM, mientras que en otros casos, como Québec (Canadá), el proceso se encuentra en marcha. En Argentina aún hay solo proyectos parlamentarios a nivel nacional, y a nivel regional hay algunas directrices como en la Provincia de Bs As a través de la Subsecretaria de Gobierno Digital que dictó las "Directrices de uso de Inteligencia Artificial Generativa en la Administración Pública de la Provincia de Buenos Aires" res. 4/2025[22] y en noviembre de 2025 la Res. 9/2025 D[23] "Reglas para el desarrollo, implementación y uso responsable de sistemas de Inteligencia Artificial para la Administración Pública de la Provincia de Buenos Aires" basada en el criterio de riesgos.

6. Conclusiones

La inteligencia artificial se ha convertido en un componente estructural de la industria de la moda y sus decisiones afectan tanto procesos productivos como representaciones simbólicas y condiciones de consumo. La aplicación de una tipología de sesgos sistémicos, estadísticos y humanos permite evidenciar que las tecnologías no son neutrales y que pueden reforzar desigualdades ya existentes. El análisis del greenwashing algorítmico mostró que los datos y los modelos pueden utilizarse para construir narrativas engañosas de sostenibilidad, lo que exige nuevas formas de supervisión y de rendición de cuentas tanto públicas como privadas.

La incorporación de la perspectiva de la discriminación algorítmica contribuye a conectar estos fenómenos con los estándares internacionales de ética en el uso de la IA poniendo como objetivo, centro y norte al ser humano y observándose que no existen soluciones desde un solo enfoque sino que requiere de estructuras, decisiones, acciones desde lo técnico y desde lo social, existiendo consenso en que las respuestas son básicamente **sociotécnicas**.

La inteligencia artificial evoluciona a un ritmo acelerado y se perfecciona de manera constante. Ese dinamismo resulta imparable y conlleva beneficios evidentes: mayor capacidad de análisis, optimización de procesos, reducción de costos y posibilidades creativas inéditas y exige inexorablemente que se desarrollen dentro de estándares mínimos de equidad, seguridad y respeto por la dignidad humana, que eviten desvíos, inequidades o prácticas abiertamente no éticas. En una industria tan simbólica como la moda que construye identidades, la forma en que se diseñan y gestionan los sistemas y modelos de IA debe estar guiada por un uso ético y responsable, alineado con principios de diversidad, transparencia, trazabilidad y control humano significativo.

Notas

1. Unesco (2023), Artículo "Lexico de la Inteligencia Artificial", En línea https://www.unesco.org/en/articles/lexicon-artificial-intelligence-0
2. Unesco (2023), Artículo "Lexico de la Inteligencia Artificial", En línea https://www.unesco.org/en/articles/lexicon-artificial-intelligence-0
3. https://dle.rae.es/sesgo
4. NIST Special Publication 1270.(2022) Towards a Standard for Identifying and Managing Bias in Artificial Intelligence d", p. ii.77. En línea https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=934464.La clasificación de sesgos sistémicos, estadísticos y humanos utilizada en este trabajo se inspira en la propuesta del National Institute of Standards and Technology (2023) sobre gestión de sesgos en sistemas de inteligencia artificial y su publicación Schwartz, Reva, Vassilev, Apóstol, Greene, Kristen, Periné, Lori, Burt, Andrew, Hall, Patrick, (2022) "
5. Schwartz, Reva, Vassilev, Apóstol, Greene, Kristen, Periné, Lori, Burt, Andrew, Hall, Patrick, (2022) "NIST Special Publication 1270. Towards a Standard for Identifying and Managing Bias in Artificial Intelligence" p 77. En línea https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=934464
6. Crawford Kate (2023) *Atlas of AI, p 332/337*
7. Joy Buolamwini (2018) Charla Ted en linea https://www-media-mit-edu.translate.goog/posts/how-i-m-fighting-bias-in-algorithms/?_x_tr_sl=auto&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=wapp
8. Comisión Europea. (2020). *Libro blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza*._https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_es
9. Unión Europea. (2019). *Directrices éticas para una inteligencia artificial fiable,* en línea https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 Ofrece directrices para su desarrollo, despliegue y utilización responsable por parte de los Estados, la industria y la sociedad civil. Respeto por los valores éticos fundamentales y consideración de sus implicancias sociales
10. Organización para la Cooperación y el Desarrollo Económico. (2019). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449). https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
11. UNESCO. (2025, febrero). ***Ética de la inteligencia artificial y justicia algorítmica****. Avances y desafíos en Ecuador y América Latina. El evento abordó el estado actual de la Readiness Assessment Methodology (RAM) en Ecuador*. **En línea** https://www.unesco.org/es/articles/etica-de-la-inteligencia-artificial-y-justicia-algoritmica
12. UNESCO (2025, octubre) Colombia avanza hacia una inteligencia artificial ética con la presentación del informe RAM de la UNESCO En linea (https://www.unesco.org/es/articles/colombia-avanza-hacia-una-inteligencia-artificial-etica-con-la-presentacion-del-informe-ram-de-la)
13. United Kingdom Government. (2023, noviembre 1-2). The Bletchley Declaration. AI Safety Summit. https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/dbc58681-1b68-47e0-8e3f-f91435fdf8ce

14. Comisión Europea. (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo de 13 de junio de 2024 sobre normas armonizadas en materia de inteligencia artificial. En linea https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689&qid=1765020718283
15. Naciones Unidas. (2024). Gobernanza de la inteligencia artificial en beneficio de la humanidad: Informe final. Pag 144 https://www par in b .un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_es.pdf
16. UNESCO. (2021). Recomendación sobre la ética de la inteligencia artificial. En línea https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa
17.UNESCO. (2021). Recomendación sobre la ética de la inteligencia artificial. En línea https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa
18. Mark D Wilkinson, Michel Dumuntier y otros (2016) The FAIR Guiding Principles for scientific data management and stewardship, Sientific Data volument 3, articulo 160018, en línea https://pmc.ncbi.nlm.nih.gov/articles/PMC4792175/ La acuñación del término FAIR Principles que corresponde al acrónico de FINDABLE (encontrable), ACESSIBLE (accesible), INTEROPRABLE (interoperable) REUSABLE se remonta a un grupo de investigadores y organizaciones del ámbito de la ciencia de datos, cuando en enero de 2014 se realizó un taller en Leiden (Países Bajos) bajo la etiqueta FORCE11/Data FAIRport" en el que se empezaron a discutir criterios para el buen manejo, reutilización y accesibilidad de datos. A partir de allí se publicó el articulo definitorio en marzo de 2016 donde se formalizan dichos cuatro principios
19. Comisión Europea. (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo de 13 de junio de 2024 sobre normas armonizadas en materia de inteligencia artificial. En linea https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689&qid=1765020718283
20. Laws/Regulations/National Standards directly regulating AI (the "AI Regulations") (2025) https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china
21. Japan's emerging framework for responsible AI: legislation, guidelines and guidance (2025) https://www.ibanet.org/japan-emerging-framework-ai-legislation-guidelines?
22. GPBA (2025) "Directrices de uso de Inteligencia Artificial Generativa en la Administración Pública de la Provincia de Buenos Aires https://normas.gba.gob.ar/documentos/Bezv5jIj.pdf
23. GPBA (2025) "Reglas para el desarrollo, implementación y uso responsable de sistemas de Inteligencia Artificial para la Administración Pública de la Provincia de Buenos Aires" P.11 https://normas.gba.gob.ar/documentos/VwRM7Ju5.pdf

Referencias bibliográficas

Comisión Europea. (2020). *Libro blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza*. https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_es

Comisión Europea. (2024). *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo* de 13 de junio de 2024 sobre normas armonizadas en materia de inteligencia artificial. En línea https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689&qid=1765020718283

Naciones Unidas. (2024). *Gobernanza de la inteligencia artificial en beneficio de la humanidad: Informe final*. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_es.pdf

Naciones Unidas. (2024, marzo). *Resolución de la Asamblea General sobre inteligencia artificial*. https://news.un.org/es/story/2024/03/1528511

OCDE. (2019). *Recomendación del Consejo sobre la Inteligencia Artificial*.
https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*.
https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

UNESCO. (2024, febrero 5-6). *Foro Global sobre la Ética de la IA*, Kranj. https://www.unesco.org/es/artificial-intelligence/recommendation-ethics?hub=32618

UNESCO. (2019). *Preliminary study on the ethics of artificial intelligence. Comisión Mundial de Ética del Conocimiento Científico y la Tecnología (COMEST)*. https://www.unesco.org/en/articles/lexicon-artificial-intelligence-0

Unión Europea. (2019). *Directrices éticas para una inteligencia artificial fiable*. https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

United Kingdom Government. (2023, noviembre 1-2). *The Bletchley Declaration*. AI Safety Summit. https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/dbc58681-1b68-47e0-8e3f-f91435fdf8ce

Buolamwini, J., y Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1–15.

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

NIST National Institute of Standards and Technology. (2023). Towards a standard for identifying and managing bias in artificial intelligence. NIST Special Publication. Schwartz, Reva, Vassilev, Apóstol, Greene, Kristen, Periné, Lori, Burt, Andrew, Hall, Patrick, (2022) "NIST Special Publication 1270. Towards a Standard for Identifying and Managing Bias in Artificial Intelligence, pag 77. En línea https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=934464

O Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.

Abstract: The incorporation of artificial intelligence (AI) systems in the design and fashion industry is profoundly transforming the value chain, from ideation and design to production, marketing, sales, logistics, and after-sales service. Its impact is increasingly profound, making strategic contributions while also presenting ethical, technical, and legal challenges, bringing sustainability, equity, and justice into convergence.
These systems are not neutral, neither in their advantages nor their risks, since the speed of their propagation can reproduce and amplify historical, social, and institutional biases. We have been hearing about "algorithmic" justice and equity for some time now.
To this end, concepts and classifications of algorithmic biases are explored, mitigation methods are examined from a sociotechnical perspective, and ethical principles, fairness, and algorithmic greenwashing are addressed.

Keywords: artificial intelligence - fashion industry - algorithmic bias - algorithmic justice - algorithmic discrimination - algorithmic greenwashing.

Resumo: A incorporação de sistemas de inteligência artificial (IA) na indústria da moda e do design está transformando profundamente a cadeia de valor, da ideação e do design à produção, marketing, vendas, logística e serviço pós-venda.
Seu impacto é cada vez mais profundo, trazendo contribuições estratégicas, ao mesmo tempo que apresenta desafios éticos, técnicos e legais, convergindo sustentabilidade, equidade e justiça.
Esses sistemas não são neutros, nem em suas vantagens nem em seus riscos, uma vez que a velocidade de sua propagação pode reproduzir e amplificar vieses históricos, sociais e institucionais. Há algum tempo ouvimos falar sobre justiça e equidade "algorítmicas".
Para tanto, são explorados conceitos e classificações de vieses algorítmicos, métodos de mitigação são examinados a partir de uma perspectiva sociotécnica e princípios éticos, justiça e greenwashing algorítmico são abordados.

Palavras-chave: inteligência artificial - indústria da moda - viés algorítmico - justiça algorítmica - discriminação algorítmica - greenwashing algorítmico.

[Las traducciones de los abstracts fueron supervisadas por el autor de cada artículo.]