CAPÍTULO 15

Discriminación algorítmica y vulneración del principio de igualdad: riesgo para colectivos vulnerables

Algorithmic discrimination and violation of the principle of equality: risks for vulnerable groups

1 **Daniel Fernando Sandoya Yance**
Magíster en Derecho Constitucional
Universidad Tecnológica Empresarial de Guayaquil, Ecuador
dsandoya@uteg.edu.ec
https://orcid.org/0009-0008-3298-1129

DOI https://doi.org/10.64644/9789907956313_15 RECIBIDO 11/09/2026 | ACEPTADO 14/10/2026

RESUMEN

La inclusión acelerada de sistemas de inteligencia artificial (IA) en los procesos de toma de decisiones públicas y privadas ha puesto en tensión el principio jurídico de igualdad, en la medida en que los algoritmos pueden reproducir, ocultar o amplificar sesgos preexistentes en los datos de entrenamiento y en las decisiones de diseño técnico. Este capítulo analiza los riesgos que la discriminación algorítmica representa para los colectivos vulnerables, mujeres, personas con discapacidad, personas migrantes, grupos étnicos y personas con procesos o antecedentes judiciales de carácter penal, a partir de una revisión integrativa cualitativa de la literatura académica publicada entre 2022 y 2025. La búsqueda documental se realizó en Dialnet, Scopus, Web of Science, Redalyc, SciELO y DOAJ, empleando operadores booleanos que combinaron los términos discriminación algorítmica, sesgo, inteligencia artificial e igualdad, y aplicando criterios de inclusión centrados en artículos revisados por pares con DOI verificable. Tras un proceso de cribado documental en cinco fases y una evaluación de calidad diferenciada, se seleccionaron quince estudios primarios, analizados mediante análisis temático inductivo. Los hallazgos evidencian que la discriminación algorítmica se manifiesta de forma directa e indirecta en el empleo, la salud, la movilidad humana, la justicia penal y el acceso a servicios, afectando desproporcionadamente a los colectivos vulnerables debido a la opacidad de los sistemas, la escasa representatividad de los datos de entrenamiento y la insuficiencia del andamiaje jurídico que regula el uso de la inteligencia artificial. La evidencia disponible es, sin embargo, predominantemente doctrinal, se concentra en contextos europeos y presta escasa atención a colectivos como las personas mayores y la infancia. Como aporte específico, el capítulo propone un esquema que relaciona los enfoques teóricos de la igualdad con los tipos y mecanismos de discriminación algorítmica, y compara los modelos regulatorios de la Unión Europea, Brasil y Colorado. Se concluye que resulta imprescindible fortalecer, de manera articulada, la transparencia técnica, la igualdad sustantiva y el desarrollo

normativo en materia de IA, con especial atención a los contextos latinoamericanos, todavía insuficientemente representados en la literatura revisada.

Palabras clave: discriminación algorítmica, principio de igualdad, transparencia, sesgos algorítmicos, derechos fundamentales

ABSTRACT

The rapid integration of artificial intelligence (AI) systems into public and private decision-making processes has placed strain on the legal principle of equality, as algorithms can reproduce, conceal, or amplify pre-existing biases found in training data and technical design choices. This chapter analyzes the risks that algorithmic discrimination poses to vulnerable groups, including women, people with disabilities, migrants, ethnic groups, and individuals with past or ongoing involvement in the criminal justice system, based on a qualitative integrative review of academic literature published between 2022 and 2025. The literature search was conducted across Dialnet, Scopus, Web of Science, Redalyc, SciELO, and DOAJ, using Boolean operators to combine the terms algorithmic discrimination, bias, artificial intelligence, and equality, and applying inclusion criteria focused on peer-reviewed, indexed articles with verifiable DOI codes. Following a five-stage screening process and a differentiated quality appraisal, fifteen primary studies were selected and analyzed through inductive thematic analysis. The findings show that algorithmic discrimination manifests both directly and indirectly in employment, healthcare, human mobility, criminal justice, and access to services, disproportionately affecting vulnerable groups due to system opacity, the underrepresentation of training data, and the inadequacy of the legal frameworks regulating AI use. However, the available evidence is predominantly doctrinal, concentrated in European contexts, and pays little attention to groups such as older persons and children. As a specific contribution, the chapter proposes a framework linking theoretical approaches to equality with the types and mechanisms of algorithmic discrimination, and compares the regulatory models of the European Union, Brazil, and Colorado. The chapter concludes that strengthening transparency, substantive equality, and regulatory development regarding AI is essential, with particular attention to Latin American contexts, which remain underrepresented in the reviewed literature.

Keywords: algorithmic discrimination, principle of equality, transparency, algorithmic biases, fundamental rights

1. Introducción

La inteligencia artificial (IA) ha experimentado un notable avance en la vida cotidiana de la sociedad contemporánea. La versatilidad de esta herramienta ha permitido acelerar el ritmo de múltiples actividades, convirtiéndose en un recurso valioso no solo para las personas, sino también para el propio Estado, que la emplea en la planificación, la elaboración de proyectos, el diseño de políticas públicas, la adopción de decisiones en el ámbito administrativo e incluso en el ámbito judicial, mediante su uso en la resolución de casos.

No obstante, junto a estos aspectos positivos —mayor eficiencia, conveniencia y, hasta cierto punto, un grado de democratización del acceso a servicios—, su empleo en determinados ámbitos podría afectar derechos fundamentales de grupos sociales y el principio de igualdad, en particular de personas migrantes y grupos étnicos, cuyos derechos se encuentran reconocidos en diversos

instrumentos internacionales. La literatura académica ha cuestionado crecientemente el uso de esta herramienta, advirtiendo que, en ocasiones, amplifica patrones históricos de desigualdad, dando lugar a lo que doctrinalmente se conoce como discriminación algorítmica (Iturmendi Rubia, 2023).

Desde el punto de vista jurídico, la inteligencia artificial ha experimentado una escalada de uso en el sistema de justicia, donde se ha convertido en una herramienta de gran utilidad, pero con escasa regulación, lo que podría estar afectando la forma en que se resuelven los casos. En este ámbito se ve amenazado el principio de igualdad y no discriminación, pese a que dicho principio constituye un pilar fundamental del derecho constitucional contemporáneo.

En primer lugar, Adams-Prassl et al. (2023) observan que la opacidad técnica de los modelos, especialmente los basados en aprendizaje automático, dificulta la identificación de las personas que han recibido un trato desigual y, con mayor razón, que estas puedan aportar los indicios y pruebas que tradicionalmente exige el derecho antidiscriminatorio. En segundo lugar, Capellà i Ricart (2024) indica que la discriminación algorítmica opera mediante variables proxy, es decir, correlaciones estadísticas con características protegidas como el sexo, el origen étnico o la discapacidad, sin que se requiera una intención discriminatoria para que se produzca un daño lesivo.

La necesidad de generar conocimiento que motiva esta revisión nace, precisamente, de la brecha existente entre el ritmo de despliegue de sistemas de IA en sectores sensibles y la capacidad de los marcos jurídicos, éticos y técnicos para detectar, anticipar y remediar sus efectos discriminatorios sobre colectivos expuestos a una vulnerabilidad constante, como las mujeres, las personas con discapacidad, las personas migrantes, los grupos étnicos, las personas adultas mayores y quienes enfrentan procesos o antecedentes judiciales. Amaya-Santos et al. (2024) señalan que estos colectivos comparten una posición de desventaja estructural previa a la introducción de la tecnología, lo que hace que los sesgos de los datos, que reflejan desigualdades históricas, tiendan a perpetuar el perjuicio cuando son empleados para entrenar modelos predictivos.

La importancia de abordar este fenómeno de forma científica y académica radica en su carácter transversal. La discriminación algorítmica no es un problema exclusivamente informático ni exclusivamente jurídico, sino un objeto de estudio inherentemente interdisciplinario que convoca también a la sociología de la tecnología y a los estudios de género y de discapacidad.

Comprender sus mecanismos de producción —sesgos de medición, de muestreo, de despliegue y algorítmicos en sentido estricto— resulta indispensable no solo para el diseño de sistemas más justos, sino también para la actualización de los marcos normativos, como señala Martínez Navarro (2025). La reciente aprobación del Reglamento (UE) 2024/1689 sobre inteligencia artificial (Parlamento Europeo y Consejo de la Unión Europea, 2024) introduce, por primera vez en el contexto europeo, un régimen específico de gestión de riesgos para los sistemas que puedan afectar a colectivos vulnerables.

Este capítulo se justifica en la constatación de que buena parte de la literatura disponible se encuentra fragmentada por disciplinas y por sectores de aplicación —salud, justicia, trabajo, educación—, sin que exista una síntesis que integre esta evidencia bajo el prisma del principio de igualdad. Esta dispersión dificulta que los responsables de políticas públicas, los operadores

jurídicos y las personas diseñadoras de sistemas cuenten con una visión de conjunto de los riesgos identificados y de los vacíos de conocimiento pendientes.

Frente a las revisiones previas sobre inteligencia artificial y discriminación, que suelen concentrarse en un único sector, en un único colectivo —como la revisión de alcance de El Morr et al. (2024) sobre discapacidad— o en una sola tradición disciplinar, técnica o jurídica, la contribución específica de este capítulo es triple. En primer lugar, integra bajo un mismo marco analítico, centrado en el principio de igualdad, evidencia procedente de estudios jurídicos, técnicos y filosóficos que hasta ahora se encuentra dispersa, y delimita con precisión las categorías conceptuales que la literatura suele emplear de forma indistinta.

En segundo lugar, propone un esquema de síntesis propio (Tabla 4) que relaciona explícitamente los enfoques teóricos de la igualdad con los tipos de discriminación algorítmica, sus mecanismos de producción y los sectores en que predominan. En tercer lugar, incorpora un análisis comparado de tres modelos regulatorios —Unión Europea, Brasil y Colorado— y examina las implicaciones de la escasez de evidencia latinoamericana para la transferibilidad de los hallazgos a la región, una dimensión escasamente abordada por las revisiones existentes.

El objetivo general de este capítulo es analizar, mediante una síntesis integrativa de la literatura académica publicada entre 2022 y 2025, los riesgos que la discriminación algorítmica plantea al principio de igualdad respecto de los colectivos vulnerables, identificando sus principales manifestaciones sectoriales, los enfoques teóricos empleados para su comprensión, los avances regulatorios y técnicos más recientes, y los vacíos de investigación que persisten en la materia.

Específicamente, se busca: i) sistematizar las tendencias temáticas de la producción científica reciente sobre discriminación algorítmica; ii) comparar los principales enfoques teóricos (anticlasificatorio, antisubordinatorio, interseccional y de justicia distributiva) empleados para explicar el fenómeno; iii) describir los avances científicos y normativos más relevantes en sectores críticos como el empleo, la salud y la justicia; y iv) señalar los vacíos y desafíos de investigación que deben orientar la agenda futura, con particular atención a los contextos latinoamericanos.

1.1. Delimitación conceptual

Dado que el fenómeno estudiado se sitúa en la intersección de categorías jurídicas, éticas, tecnológicas y sociológicas, resulta necesario precisar desde el inicio el alcance de cuatro nociones que la literatura suele emplear de forma intercambiable y que, en este capítulo, se utilizan con un sentido diferenciado.

El *sesgo algorítmico* se entiende como una desviación sistemática en el funcionamiento o en los resultados de un sistema automatizado, originada en los datos de entrenamiento (sesgos de muestreo o de medición), en las decisiones de diseño (selección de variables, definición del objetivo de optimización) o en las condiciones de despliegue e interacción con las personas usuarias; Pérez-Ugena Coromina, 2024). Se trata de una categoría técnica y descriptiva: no todo sesgo es jurídicamente relevante ni produce, por sí mismo, un trato discriminatorio.

La *inequidad estadística*, o disparidad de resultados, designa la diferencia cuantificable que un sistema produce entre grupos en las tasas de acierto, error, selección o asignación de recursos, medida mediante métricas de equidad como la paridad demográfica o la igualdad de tasas de falsos positivos y falsos negativos (Calahorrano Latorre, 2025). Es una noción operativa: constituye un indicio, pero no una prueba concluyente, de discriminación, ya que determinadas disparidades pueden estar justificadas y, a la inversa, la paridad estadística no excluye por sí sola un trato lesivo.

La *discriminación algorítmica* es, en cambio, una categoría jurídico-normativa: se configura cuando el uso de un sistema automatizado produce un trato menos favorable, o una desventaja particular, a personas en razón de una característica protegida, ya sea de forma directa, cuando el sistema utiliza esa característica o una variable sustitutiva que la reproduce con precisión, o indirecta, cuando un criterio aparentemente neutro genera un impacto desproporcionado sin justificación objetiva y proporcional, con independencia de la intención de quienes diseñan o despliegan el sistema (Adams-Prassl et al., 2023; Lousada Arochena, 2024).

Por último, la *vulneración del principio de igualdad* se sitúa en el plano constitucional y de los derechos fundamentales: alude a la afectación del mandato de igualdad formal y sustantiva y de la prohibición de discriminación, que puede derivar no solo de decisiones individuales discriminatorias, sino también de la acumulación de efectos estructurales que consolidan la subordinación de determinados grupos (Capellà i Ricart, 2024).

En consecuencia, el capítulo asume entre estas categorías una relación de progresión analítica, y no de sinonimia: el sesgo algorítmico es el mecanismo técnico; la inequidad estadística, su expresión medible; la discriminación algorítmica, su calificación jurídica cuando concurren los elementos exigidos por el derecho antidiscriminatorio; y la vulneración del principio de igualdad, su consecuencia en el plano de los derechos fundamentales. Esta distinción orienta la lectura de los resultados y permite identificar en qué nivel se sitúa cada hallazgo de la literatura revisada.

2. Metodología de la revisión

El presente trabajo se elabora a partir de una revisión integrativa de la literatura académica. Se optó por este tipo de diseño, y no por una revisión sistemática en sentido estricto ni por un estado del arte meramente descriptivo, en la medida en que el fenómeno de la discriminación algorítmica ha sido abordado desde tradiciones metodológicas heterogéneas —estudios doctrinales y dogmático-jurídicos, investigaciones empíricas cuantitativas sobre sesgo en modelos de aprendizaje automático, y ensayos filosóficos y reflexiones normativas— que difícilmente podrían compararse bajo los criterios homogéneos de calidad metodológica que exige una revisión sistemática clásica. La revisión integrativa permite combinar evidencia proveniente de diseños diversos con el fin de generar una comprensión más global del fenómeno, sin renunciar al rigor ni a la transparencia en la estrategia de búsqueda y selección documental.

Conviene precisar, para evitar ambigüedades sobre el tipo de revisión desarrollada, que la adopción de elementos procedimentales propios de las revisiones sistemáticas, búsqueda

estructurada en bases de datos, criterios explícitos de inclusión y exclusión, cribado por fases y evaluación de calidad, no convierte este trabajo en una revisión sistemática.

Dichos elementos se emplean como garantías de transparencia y trazabilidad de la búsqueda, mientras que la síntesis de la evidencia es de naturaleza cualitativa e interpretativa, como corresponde al diseño integrativo, cuyas etapas, identificación del problema, búsqueda bibliográfica, evaluación de los datos, análisis de los datos y presentación de los resultados, se siguieron en este estudio. Del mismo modo, la declaración PRISMA 2020, Page et al. (2021) se utilizó únicamente como guía para reportar el proceso de selección documental (Tabla 1 y Figura 1), y no como protocolo de una revisión sistemática con metanálisis.

La búsqueda documental se llevó a cabo entre los meses de julio y agosto de 2025, restringiendo los resultados a documentos publicados entre enero de 2022 y agosto de 2025, con el fin de garantizar la actualidad de la evidencia analizada. Las cadenas de búsqueda se ajustaron a la sintaxis específica de cada base de datos y se reforzaron mediante una revisión manual de las referencias bibliográficas de los documentos preseleccionados (técnica de bola de nieve), con el fin de identificar estudios pertinentes que no estuvieran indexados directamente bajo los descriptores utilizados. Las bases y repositorios consultados fueron Scopus, Web of Science, Scielo, Redalyc, Dialnet y DOAJ (Directory of Open Access Journals).

En cuanto a los criterios de inclusión se determinaron:

- Artículos publicados en revistas académicas indexadas y sometidos a revisión por pares.
- Documentos publicados entre enero de 2022 y agosto de 2025.
- Estudios que abordaran de manera explícita la discriminación algorítmica, los sesgos algorítmicos o el principio de igualdad en relación con sistemas de inteligencia artificial.
- Textos disponibles en español, inglés o portugués, con resumen y texto completo accesibles.
- Documentos que contaran con código DOI verificable, como garantía mínima de trazabilidad y calidad editorial.

En cuanto a los criterios de exclusión se determinaron mediante:

- Documentos duplicados entre las distintas bases de datos consultadas.
- Literatura no académica (blogs, notas de prensa, entradas de opinión sin respaldo editorial revisado por pares).
- Estudios sin descripción metodológica mínima que permitiera valorar el origen y la fiabilidad de sus hallazgos.
- Documentos sin código DOI verificable o cuya autoría no pudo confirmarse.
- Publicaciones anteriores a 2022, salvo instrumentos normativos vigentes citados como fuente primaria de derecho.

2.1. Procedimiento de selección

El proceso de selección documental se reporta siguiendo la estructura del diagrama de flujo PRISMA 2020 (Page et al., 2021), adaptada a las características de una revisión integrativa que combina literatura jurídica, técnica y empírica. La Tabla 1 sintetiza el número de registros identificados, cribados, evaluados a texto completo e incluidos en la síntesis final, y la Figura 1 representa gráficamente dicho flujo.

Tabla 1

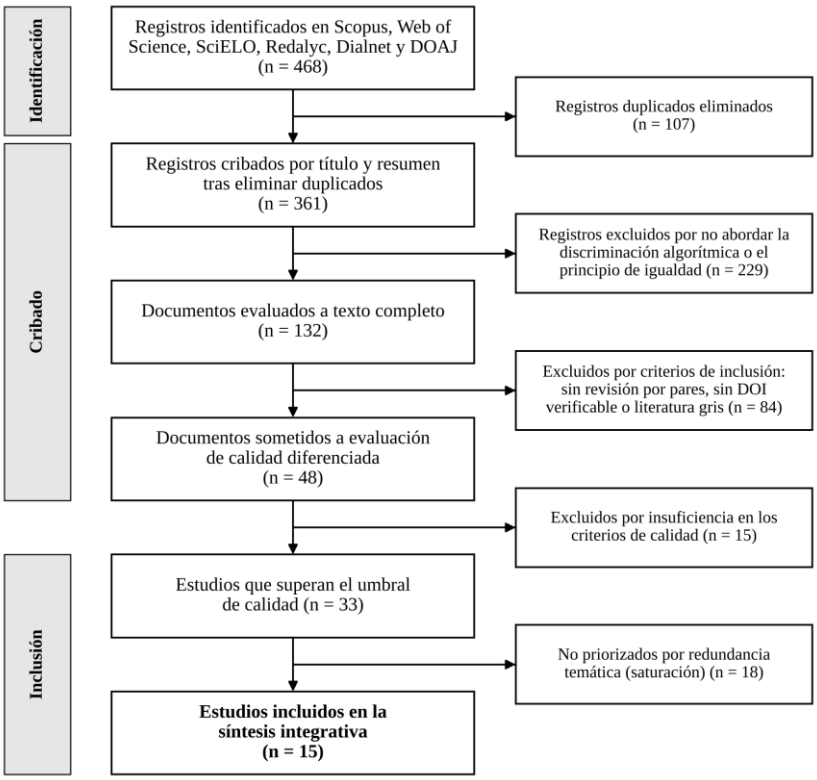
Proceso de identificación, cribado y selección documental

Fase del proceso	Registros	Observaciones
Identificación	n = 468	Registros localizados en las seis bases de datos consultadas mediante las cadenas de búsqueda booleanas.
Eliminación de duplicados	n = 361	Se depuraron 107 registros duplicados entre bases de datos.
Cribado por título y resumen	n = 132	Se excluyeron 229 registros por no abordar directamente la discriminación algorítmica o el principio de igualdad.
Evaluación de texto completo	n = 48	Se descartaron 84 documentos por no cumplir los criterios de inclusión (ausencia de revisión por pares, falta de DOI verificable, literatura gris).
Evaluación de calidad	n = 33	Se excluyeron 15 documentos adicionales por insuficiencia en los criterios de calidad diferenciada descritos en el apartado “Criterios de calidad de los estudios incluidos” (descripción metodológica incompleta, ausencia de fundamentación teórica explícita o redundancia temática sin aporte conceptual adicional).
Estudios incluidos en la síntesis	**n = 15**	De los 33 estudios que superaron el umbral de calidad, se priorizaron 15 conforme a los criterios de saturación temática y representatividad teórica; los 18 restantes no se incorporaron por reiterar casos o argumentos ya representados en el corpus.

Nota. Elaboración propia a partir del proceso de búsqueda y cribado documental descrito.

Figura 1

Diagrama de flujo del proceso de selección documental



Nota. Elaboración propia, adaptada de la estructura del diagrama de flujo PRISMA 2020 (Page et al., 2021), a partir de los datos de la Tabla 1.

La reducción del corpus de 48 a 15 estudios no respondió a un límite arbitrario, sino a la aplicación sucesiva de los criterios de calidad y de saturación temática. En un primer momento se identificaron 33 estudios que superaban el umbral mínimo de calidad exigido; de ellos, se priorizaron los quince que, en conjunto, permitían representar los principales sectores de aplicación identificados en la literatura (empleo, salud, justicia penal y migración), los cuatro enfoques teóricos que estructuran el debate (anticlasificatorio, antisubordinatorio, interseccional y de justicia distributiva) y la mayor diversidad geográfica y metodológica posible dentro de las limitaciones de la producción disponible.

Estudios adicionales que abordaban el mismo caso paradigmático (por ejemplo, distintos análisis doctrinales sobre el sistema COMPAS o sobre el mismo caso de cribado curricular sesgado por género) fueron excluidos cuando no aportaban elementos teóricos o empíricos sustancialmente nuevos respecto de los ya incorporados, con el fin de evitar la sobrerrepresentación de un mismo caso en desmedro de la diversidad sectorial y teórica del corpus.

Esta decisión se apoya en el criterio de saturación temática propio de la investigación cualitativa, entendido aquí, en su dimensión temática, como el punto en que la incorporación de nuevos estudios a la matriz de vaciado deja de generar códigos o categorías adicionales y se limita a reiterar los ya identificados. Bajo este criterio, la representatividad que se persigue es teórica y no estadística: el corpus no pretende ser una muestra proporcional de la producción científica sobre

discriminación algorítmica, sino un conjunto suficiente para cubrir la variedad de enfoques teóricos, mecanismos de sesgo y sectores de aplicación presentes en la literatura. Este planteamiento es coherente con la finalidad interpretativa de la revisión integrativa, pero impide extraer conclusiones sobre la frecuencia o la prevalencia de los fenómenos descritos.

Debe reconocerse, asimismo, que la composición final del corpus presenta una marcada concentración geográfica. Siete de los estudios incluidos proceden de España y uno del Reino Unido, de modo que más de la mitad del corpus se sitúa en el contexto europeo; otros dos corresponden a Estados Unidos y Canadá, y solo cuatro tienen origen latinoamericano (Colombia, Ecuador y México), dos de los cuales se centran, además, en marcos normativos o casos extranjeros (Tabla 2).

Esta distribución refleja en parte la estructura de la producción indexada disponible en el periodo y en los idiomas considerados, pero condiciona el alcance de los hallazgos: las categorías, los casos y los marcos regulatorios analizados responden mayoritariamente a contextos europeos y estadounidenses, por lo que su transferencia a otras realidades debe realizarse con cautela, cuestión que se retoma en la discusión.

2.2. Criterios de calidad de los estudios incluidos

Dado que el corpus documental combina tradiciones metodológicas heterogéneas, estudios doctrinal-jurídicos, empírico-cualitativos, empírico-documentales, filosófico-antropológicos y analítico-normativos, no resultaba adecuado aplicar una única herramienta estandarizada de evaluación de calidad, como ocurriría en una revisión sistemática centrada en un solo tipo de diseño. En su lugar, se adoptó un esquema de evaluación diferenciada de calidad, adaptado de los principios generales de las listas de verificación del tipo Joanna Briggs Institute (JBI) y CASP, aplicado de forma distinta según la naturaleza del estudio:

- Para los estudios de corte empírico (cualitativo, documental o cuantitativo): claridad de la pregunta de investigación, adecuación y transparencia de la descripción metodológica, coherencia entre los datos presentados y las conclusiones alcanzadas, y explicitación de las limitaciones del propio estudio.
- Para los estudios doctrinal-jurídicos y dogmático-jurídicos: actualidad y suficiencia de las fuentes normativas y jurisprudenciales empleadas, coherencia argumentativa interna, diálogo explícito con posiciones doctrinales divergentes y aporte a la clarificación conceptual del fenómeno analizado.
- Para los ensayos filosóficos y analítico-normativos: coherencia argumentativa, vinculación explícita con los debates teóricos existentes sobre igualdad y no discriminación, y originalidad conceptual del planteamiento.

Cada uno de los 48 estudios evaluados a texto completo fue valorado de forma independiente frente a estos criterios, asignando una valoración cualitativa (adecuado / parcialmente adecuado / inadecuado) por dimensión. Los estudios valorados como inadecuados en dos o más dimensiones fueron excluidos del corpus final; los resultados de esta evaluación no se emplearon para ponderar

el peso argumentativo de cada estudio dentro de la síntesis, sino únicamente como filtro de inclusión, dado que la revisión integrativa no persigue un metanálisis cuantitativo de efectos.

2.3. Procedimiento de análisis temático

El análisis del corpus documental siguió las fases del análisis temático propuestas por Braun y Clarke (2021), adaptadas al carácter interdisciplinario del material: i) familiarización con los documentos mediante una lectura completa de cada estudio; ii) generación de códigos descriptivos iniciales, aplicados a los fragmentos referidos a mecanismos de sesgo, colectivos afectados, enfoques teóricos, sectores de aplicación y respuestas regulatorias; iii) agrupación de los códigos en categorías temáticas emergentes; iv) revisión y refinamiento de las categorías mediante contraste constante con el corpus completo, verificando su coherencia interna y su capacidad de diferenciación externa; v) definición y denominación final de los temas, que dieron lugar a la estructura de los apartados, y vi) elaboración de una matriz de vaciado documental que registró, para cada estudio, el enfoque teórico predominante, el colectivo o colectivos abordados, el sector de aplicación, el país o región de origen, los principales hallazgos y las citas textuales relevantes. Esta matriz constituye el soporte analítico que permitió construir tanto la Tabla 2 como los apartados de desarrollo y discusión del capítulo.

Finalmente, con el fin de precisar el marco conceptual y de contrastar los hallazgos del corpus con evidencia empírica reciente, en la introducción y en la discusión se recurre a literatura complementaria, trabajos metodológicos, estudios empíricos de auditoría algorítmica y estudios latinoamericanos, que no forma parte del corpus analizado, por lo que no se incluye en la Tabla 2 ni intervino en la construcción de las categorías temáticas.

2.4. Limitaciones metodológicas de la revisión

Como toda revisión integrativa que combina literatura de naturalezas metodológicas distintas, este trabajo presenta limitaciones que deben explicitarse. En primer lugar, la evaluación de calidad diferenciada, si bien permite comparar estudios de distinta naturaleza, conlleva un mayor grado de juicio interpretativo que las herramientas estandarizadas de evaluación del riesgo de sesgo propias de las revisiones sistemáticas clásicas, lo que introduce un margen de subjetividad que no puede eliminarse por completo.

En segundo lugar, la restricción idiomática a español, inglés y portugués, junto con la exigencia de DOI verificable, pudo excluir literatura relevante publicada en otros idiomas o en repositorios institucionales no indexados, lo que probablemente contribuye al sesgo de concentración geográfica identificado. En tercer lugar, el tamaño reducido del corpus final (n = 15), aunque coherente con el propósito de una síntesis integrativa en profundidad, limita la posibilidad de generalizar los hallazgos y no permite un análisis cuantitativo de tendencias. En cuarto lugar, el corpus está dominado por estudios de naturaleza doctrinal, teórica o filosófica: solo tres de los trabajos incluidos tienen un componente empírico o de síntesis de evidencia empírica (Amaya-Santos et al., 2024; El Morr et al., 2024; Roa Avella et al., 2022), y ninguno consiste en una auditoría directa de

sistemas desplegados en contextos reales, lo que condiciona el tipo de conclusiones que pueden extraerse.

Finalmente, al no tratarse de una revisión sistemática en sentido estricto, no se registró un protocolo previo en una plataforma de registro de revisiones, y el diagrama de flujo de la Figura 1 se elaboró a partir del registro del proceso de cribado; ello se reconoce como una limitación de transparencia metodológica frente a los estándares más exigentes del campo, que futuras actualizaciones de esta línea de investigación deberían subsanar.

Aplicados los criterios de inclusión, exclusión y calidad, el corpus final quedó constituido por quince estudios primarios, cuya distribución por base de datos, país de origen, enfoque metodológico y colectivo abordado se presenta en la Tabla 2. Esta recopilación fue analizada mediante la matriz temática que permitió identificar categorías emergentes, tendencias de investigación, enfoques teóricos, avances sectoriales y vacíos de conocimiento, los cuales estructuran el desarrollo de la revisión en el apartado siguiente.

Tabla 2

Corpus documental analizado (n = 15)

Estudio	Base de datos	País	Enfoque metodológico	Colectivo abordado
Adams-Prassl et al. (2023)	Web of Science	Reino Unido	Doctrinal-jurídico	Trabajadores en general
Álvarez García (2023)	Dialnet	España	Doctrinal-jurídico	Personas con discapacidad
Amaya-Santos et al. (2024)	SciELO	España	Empírico-cualitativo	Pacientes / salud
Calahorrano Latorre (2025)	DOAJ	Ecuador	Doctrinal-dogmático	Pacientes / salud
Capellà i Ricart (2024)	DOAJ	España	Teórico-jurídico	Grupos subordinados estructuralmente
El Morr et al. (2024)	Scopus	Canadá	Revisión sistemática exploratoria	Personas con discapacidad
Flores Anarte (2023)	Dialnet	España	Doctrinal-jurídico	Mujeres
Iturmendi Rubia (2023)	Redalyc	España	Doctrinal-jurídico	Personas migrantes
Lousada Arochena (2024)	Dialnet	España	Doctrinal-jurídico	Mujeres
Martínez Navarro (2025)	Redalyc	Colombia / UE	Analítico-normativo	Colectivos vulnerables (general)
Morales Ramírez (2023)	Redalyc	México	Filosófico-antropológico	Población general
Pérez-Ugena Coromina (2024)	Dialnet	España	Doctrinal-jurídico	Mujeres
Roa Avella et al. (2022)	Dialnet	Colombia / EE. UU.	Empírico-documental	Personas afrodescendientes / justicia penal
Tilmes (2022)	Scopus	EE. UU.	Ético-filosófico	Personas con discapacidad

Nota. Elaboración propia a partir del corpus documental seleccionado.

3. Desarrollo de la revisión

El análisis del corpus documental permite identificar una primera tendencia: el desplazamiento de la discusión académica desde una preocupación inicial por la discriminación algorítmica como fenómeno técnico vinculado a la calidad de los datos de entrenamiento, hacia una comprensión más amplia que integra sus dimensiones jurídicas, éticas y de derechos humanos. Iturmendi Rubia (2023) expone cómo la discriminación algorítmica, entendida como el fenómeno por el cual los sistemas de aprendizaje automático perpetúan o amplifican prejuicios humanos preexistentes, ha pasado de analizarse como un defecto técnico corregible a considerarse un riesgo estructural para la dignidad humana, particularmente en el caso de las personas migrantes, quienes suelen estar sobreexpuestas en bases de datos de control fronterizo y de seguridad sin las garantías procesales adecuadas.

La segunda tendencia identificada es una creciente segmentación de la investigación. En vez de estudios generales sobre "sesgo en IA", la producción reciente tiende a estudiarse por ámbito de aplicación: empleo y selección de personal (Adams-Prassl et al., 2023; Tilmes, 2022), salud y diagnóstico clínico (Amaya-Santos et al., 2024; Calahorrano Latorre, 2025), justicia penal (Roa Avella et al., 2022) y protección social o migratoria (Iturmendi Rubia, 2023).

Esta división responde a la constatación de que los mecanismos de producción del sesgo, así como sus consecuencias jurídicas, varían esencialmente según el contexto de aplicación: no es lo mismo un sesgo en un algoritmo de cribado curricular que uno incorporado en un sistema de valoración del riesgo de reincidencia penal, aun cuando ambos puedan analizarse bajo la sombra común del principio de igualdad.

La entrada en vigor del Reglamento (UE) 2024/1689 el 1 de agosto de 2024 (Parlamento Europeo y Consejo de la Unión Europea, 2024), analizado por Martínez Navarro (2025), ha generado una oleada de análisis jurídicos orientados a evaluar su capacidad real para prevenir la discriminación algorítmica, especialmente respecto de los colectivos vulnerables mencionados de forma expresa en la norma. Esta inclinación normativa convive con una llamativa insuficiencia de estudios empíricos latinoamericanos que examinen la aplicación de marcos regulatorios equivalentes dentro de la región, lo que sugiere una desproporción geográfica en la producción científica.

Finalmente, se observa una cuarta tendencia relativa al tránsito desde enfoques puramente correctivos, centrados en la auditoría y mitigación técnica del sesgo una vez detectado, hacia enfoques preventivos y de diseño que buscan incorporar la perspectiva de género desde las primeras fases del desarrollo, según Pérez-Ugena Coromina (2024). Para El Morr et al. (2024), la inteligencia artificial aporta a las personas con discapacidad ventajas esenciales que se concentran en tres ejes: facilitar el acceso a entornos y servicios, fortalecer su independencia en la vida cotidiana y promover una mayor participación e inclusión social. Este giro hacia el diseño responsable dialoga con los principios de "privacidad desde el diseño" y "ética desde el diseño", ya consolidados en otros ámbitos de la regulación tecnológica.

La bibliografía revisada permite discernir al menos cuatro enfoques teóricos, no necesariamente excluyentes entre sí, que orientan la comprensión académica de la discriminación algorítmica y su relación con el principio de igualdad. La Tabla 3 sintetiza sus premisas centrales y los autores representativos dentro de la recopilación analizada.

Tabla 3

Enfoques teóricos identificados en la revisión

Enfoque teórico	Premisa central	Autores representativos
Anticlasificación	La igualdad exige neutralidad formal: prohíbe el uso de categorías protegidas (sexo, raza, discapacidad) como criterio de decisión, incluso de forma indirecta.	Adams-Prassl et al. (2023)
Antisubordinación	La igualdad exige atender los efectos estructurales de la discriminación sobre grupos históricamente subordinados, más allá de la neutralidad formal del algoritmo.	Capellà i Ricart (2024)
Interseccional	Los sesgos algorítmicos se potencian cuando convergen varias categorías de vulnerabilidad (por ejemplo, mujer, migrante y de bajos ingresos), generando discriminación compuesta.	Amaya-Santos et al. (2024); Iturmendi Rubia (2023)
Justicia distributiva y equidad técnica	La equidad se mide mediante métricas estadísticas de paridad entre grupos en los resultados del sistema (falsos positivos, tasas de error).	Calahorrano Latorre (2025); El Morr et al. (2024)

Nota. Elaboración propia a partir del análisis temático del corpus documental.

Con el fin de visualizar de manera integrada la relación entre estos enfoques teóricos, los mecanismos de producción del sesgo y las manifestaciones sectoriales, la Tabla 4 ofrece un esquema de síntesis que constituye uno de los aportes propios de esta revisión.

Tabla 4

Relación entre enfoques teóricos, tipos de discriminación algorítmica y sectores de aplicación

Enfoque teórico dominante	Tipo de discriminación algorítmica asociada	Mecanismo de producción típico	Sector(es) donde predomina en el corpus
Anticlasificación	Discriminación directa por variable proxy	Sesgo de medición / uso de variables sustitutas de una categoría protegida	Empleo y selección de personal
Antisubordinación	Discriminación estructural / sistémica	Monocultura algorítmica; descontextualización histórica de los datos	Justicia penal; regulación general
Interseccional	Discriminación compuesta o agravada	Sesgo de muestreo combinado en múltiples ejes de identidad	Salud; protección migratoria
Justicia distributiva y equidad técnica	Disparidad estadística entre subgrupos	Sesgo de despliegue; desbalance de tasas de error entre grupos	Salud; discapacidad

Nota. Los sectores indicados corresponden a aquellos en que cada enfoque predomina dentro del corpus, no a su ámbito exclusivo de aplicación. Elaboración propia a partir del análisis temático del corpus documental.

El planteamiento anticlasificatorio, de corte liberal clásico, sostiene que la igualdad exige neutralidad formal de las decisiones respecto de categorías protegidas. Adams-Prassl et al. (2023) aplican esta lógica al analizar cómo los algoritmos pueden constituir discriminación directa cuando actúan como "proxies perfectos" de una característica protegida, esto es, cuando el conjunto de variables empleadas por el sistema reproduce con precisión la pertenencia a un grupo protegido, incluso si ningún ser humano ha tenido la intención de discriminar. Con esta visión, la responsabilidad jurídica no depende de la intencionalidad de quien programa el sistema, sino del efecto discriminatorio verificable del sistema.

El enfoque antisubordinatorio, desarrollado con claridad por Capellà i Ricart (2024), cuestiona que la neutralidad formal sea suficiente para garantizar una igualdad sustantiva. Desde esta perspectiva, la discriminación algorítmica debe analizarse atendiendo a si el sistema contribuye a prolongar relaciones estructurales de subordinación entre grupos sociales, con independencia de que la variable protegida se utilice de forma directa o indirecta.

La crítica antisubordinatoria identifica cuatro rasgos específicos: i) la incorporación de estereotipos en los conjuntos de datos de entrenamiento; ii) la apariencia engañosa de neutralidad técnica que oculta decisiones valorativas; iii) la descontextualización de los datos históricos respecto de su origen social; y iv) la tendencia a una "monocultura algorítmica", es decir, la convergencia de distintos sistemas hacia patrones de decisión similares que amplifican, a escala social, sesgos que individualmente podrían parecer menores.

El planteamiento interseccional resulta especialmente relevante para comprender la situación de los colectivos que acumulan más de una condición de vulnerabilidad. Amaya-Santos et al. (2024) plantean la necesidad de incorporar un enfoque interseccional en el diseño, entrenamiento y evaluación de sistemas de IA en salud: i) denominan "sesgo interseccional" a la discriminación sistemática e injusta que surge cuando interactúan varios ejes de identidad y opresión, generando disparidades que no son la simple suma de discriminaciones separadas, sino una forma específica y agravada de exclusión; ii) advierten que los análisis de equidad basados en una sola variable (solo género, solo etnia, etc.) son insuficientes para captar la naturaleza entrelazada de la discriminación algorítmica, lo que deja fuera a las poblaciones en los márgenes de varias categorías; y iii) proponen que, para evitar la exclusión de poblaciones vulnerables en la salud digital, es indispensable adoptar marcos analíticos y métricas que consideren de manera conjunta género, raza o etnia, clase, estatus migratorio y otras categorías relevantes.

En la misma línea, Iturmendi Rubia (2023) sostiene que ciertos colectivos migrantes enfrentan una doble vulnerabilidad algorítmica: se encuentran sobrerrepresentados en las bases de datos vinculadas con la seguridad y el control, mientras están insuficientemente representados en conjuntos de datos considerados "benignos", como aquellos utilizados para evaluar solicitudes de crédito. Esta situación puede provocar que los sistemas automatizados asocien injustificadamente la condición migratoria con el riesgo o la sospecha, mientras limitan simultáneamente el acceso de estas personas a oportunidades económicas y sociales.

El desarrollo analítico del enfoque interseccional exige, sin embargo, ir más allá de su reconocimiento declarativo. Para mostrar que el derecho antidiscriminatorio, al examinar por

separado el sexo y la raza, dejaba sin protección la experiencia específica de las mujeres negras, adquiere en el contexto algorítmico una dimensión técnica precisa: la *discriminación compuesta* se produce cuando el perjuicio que sufre un subgrupo definido por la intersección de dos o más características es mayor que el observado para cada una de ellas por separado, o incluso cuando existe en el subgrupo pese a no detectarse en ninguna de las categorías consideradas aisladamente.

Estudios empíricos recientes, ajenos al corpus, pero útiles para contrastarlo, confirman esta lógica. Wilson y Caliskan (2024) hallaron que los sistemas favorecieron nombres asociados a personas blancas en el 85,1 % de las comparaciones, mientras que los nombres asociados a mujeres fueron preferidos únicamente en el 11,1 %.

Asimismo, los hombres negros fueron desfavorecidos en hasta el 100 % de los casos, evidenciando sesgos interseccionales que no se explican por la simple adición de los sesgos de raza y género. An et al. (2025), analizan que, de aproximadamente 361 000 currículos evaluados por cinco modelos de lenguaje de gran escala, se observaron puntuaciones más favorables para las mujeres y menos favorables para los hombres negros con calificaciones equivalentes. Este resultado evidencia que la dirección del sesgo puede variar según el eje analizado y que solo se vuelve visible cuando los datos se desagregan por subgrupos interseccionales.

Desde el punto de vista de la medición, ello tiene consecuencias directas sobre las métricas de equidad. Las métricas convencionales evalúan la paridad respecto de una única variable protegida, de modo que un sistema puede satisfacer simultáneamente la paridad por sexo y la paridad por origen étnico y, sin embargo, perjudicar sistemáticamente a un subgrupo concreto, fenómeno que la literatura técnica ha denominado *fairness gerrymandering (Kearns et al., 2018).*

Para afrontarlo se han propuesto métricas de equidad interseccional, como la equidad por subgrupos, que exige que las tasas de error o de selección no difieran significativamente en ningún subgrupo definido por combinaciones de atributos protegidos (Kearns et al., 2018), o la equidad diferencial, que acota la razón entre las probabilidades de obtener un resultado favorable en todos los subgrupos interseccionales (Foulds et al., 2020).

Su aplicación enfrenta, no obstante, dos dificultades que conectan con los hallazgos del corpus: la escasez de datos sobre subgrupos pequeños, que hace estadísticamente inestables las estimaciones precisamente para las poblaciones más marginadas (Amaya-Santos et al., 2024), y la indisponibilidad o el carácter sensible de los datos sobre categorías protegidas, cuya recolección está limitada en muchos ordenamientos. De ahí que el enfoque interseccional no pueda reducirse a una cuestión métrica y requiera también un reconocimiento jurídico de la discriminación múltiple o interseccional, todavía incipiente en los marcos normativos analizados.

En la literatura de orientación más cuantitativa predomina un enfoque de justicia distributiva y equidad técnica que opera el principio de igualdad mediante indicadores estadísticos de paridad entre grupos, como las tasas comparables de falsos positivos o falsos negativos en distintas subpoblaciones.

Calahorrano Latorre (2025) y El Morr et al. (2024) sostienen que, aunque estas métricas son herramientas relevantes para identificar y medir desigualdades objetivamente verificables, resultan

insuficientes para revelar por sí mismas los efectos cualitativos de la discriminación algorítmica, entre ellos el deterioro de la confianza de las personas con discapacidad. Ese deterioro se evidencia en los sistemas de apoyo clínico y la carga simbólica derivada de que un algoritmo las etiquete reiteradamente como casos "atípicos", debido a que fue desarrollado principalmente con datos correspondientes a personas sin discapacidad o consideradas neurotípicas. La falta de representatividad de los datos puede provocar que los sistemas de inteligencia artificial ignoren a las personas con discapacidad, generando resultados discriminatorios.

3.1. Discriminación algorítmica por razón de género

La investigación sobre el sesgo de género en la inteligencia artificial representa una de las áreas de estudio más consolidadas dentro de la literatura analizada. Flores Anarte (2023) estudia la manera en que los sistemas de IA pueden reproducir los estereotipos de género incorporados en los datos utilizados para su entrenamiento, exponiendo casos relacionados con la traducción automática y el reconocimiento de voz, en los que se atribuyen sistemáticamente determinadas profesiones en función del género gramatical, aun cuando la lengua de origen no contenga marcas explícitas de género. El trabajo explica la existencia de discriminación algorítmica por razón de sexo cuando el funcionamiento del sistema ocasiona a las mujeres un perjuicio significativo y reiterado, carente de una justificación objetiva y proporcional, que afecta derechos como el acceso al empleo.

Complementando lo expuesto sobre el sesgo algorítmico de género, Pérez-Ugena Coromina (2024) identifica tres vías principales mediante las cuales puede generarse dicho sesgo: la primera corresponde al sesgo de muestreo, originado por la representación desigual de hombres y mujeres en los datos utilizados para entrenar el sistema; la segunda consiste en la reproducción de prejuicios preexistentes, que tiene lugar cuando el algoritmo aprende y replica asociaciones discriminatorias arraigadas en la sociedad; y la tercera es el sesgo de interacción, producido cuando el uso del sistema por parte de las personas usuarias contribuye a retroalimentar y consolidar el patrón discriminatorio inicial. La autora destaca que la perspectiva de género debe integrarse desde las primeras etapas de diseño del sistema, en lugar de incorporarse únicamente como una medida correctiva posterior a su implementación.

Lousada Arochena (2024) plantea si la discriminación algorítmica requiere una revisión de las categorías tradicionales del derecho antidiscriminatorio, a partir de un caso ya consolidado en la literatura sobre un sistema de selección de personal desarrollado con currículos históricos de una empresa tecnológica en los que predominaban los hombres entre las personas contratadas para cargos técnicos. A partir de esos datos, el algoritmo dedujo que el sexo masculino constituía una característica implícitamente vinculada con el desempeño de tales funciones y, en consecuencia, sancionó de manera sistemática los currículos que contenían referencias a organizaciones o actividades relacionadas con mujeres. Este ejemplo ilustra cómo un sistema algorítmico puede generar discriminación indirecta por razón de sexo, aun cuando ninguna variable relativa al género se encuentre incorporada expresamente en sus parámetros.

3.2. Discriminación algorítmica por razón de discapacidad

De acuerdo con la bibliografía examinada, las personas con discapacidad muestran una vulnerabilidad algorítmica diferenciada, producto de la concurrencia de dos elementos: (i) la insuficiente inclusión de la diversidad funcional en los conjuntos de datos de entrenamiento y (ii) la propensión de múltiples sistemas a funcionar con un estándar implícito de "cuerpo capacitado", de modo que toda divergencia respecto de ese patrón es leída como una anomalía estadística.

Álvarez García (2023) sostiene que esta doble condición sitúa a las personas con discapacidad en un escenario de exclusión agravada: de una parte, la brecha digital generada por la inaccesibilidad de numerosas interfaces tecnológicas; de otra, el sesgo algorítmico en sentido estricto, visible en sistemas de reconocimiento facial que cometen errores con mayor frecuencia al interpretar expresiones faciales atípicas, así como en aplicaciones de procesamiento del lenguaje natural que clasifican como "tóxicos" contenidos que únicamente aluden a la discapacidad.

Tilmes (2022) profundiza en el ámbito concreto de la contratación laboral, sosteniendo que las estrategias de equidad algorítmica fundadas únicamente en la paridad estadística entre grupos resultan insuficientes cuando se trata de personas con discapacidad, dado que esta no es una categoría binaria ni homogénea, sino un espectro de condiciones marcadamente heterogéneo. Frente a ello, el autor propone reforzar los mecanismos técnicos de mitigación del sesgo con ajustes razonables individualizados, trasladando al entorno algorítmico un principio ya consolidado en el derecho antidiscriminatorio en materia de discapacidad.

En la misma línea, El Morr et al. (2024) proponen adoptar el modelo social de la discapacidad como fundamento del desarrollo de la inteligencia artificial. Desde esta perspectiva, la discapacidad no constituye una deficiencia inherente a la persona, sino que surge de la interacción entre esta y un entorno, incluidos los sistemas algorítmicos, que no ha sido adecuadamente adaptado a sus necesidades. El énfasis recae, en consecuencia, en soluciones de carácter técnico, como interfaces adaptadas, sistemas de reconocimiento o dispositivos de apoyo, orientadas a eliminar barreras de uso en entornos digitales y físicos.

3.3. Discriminación algorítmica en la justicia penal

La justicia penal se ha convertido en un escenario central del debate sobre discriminación algorítmica, en buena medida por la polémica generada por el sistema COMPAS, utilizado en varios estados de Estados Unidos para estimar la probabilidad de reincidencia delictiva. Roa Avella et al. (2022) documentan que las investigaciones periodísticas y académicas encontraron que, entre las personas que no reincidieron, las afrodescendientes tenían una probabilidad significativamente mayor que las blancas de haber sido clasificadas erróneamente como de "alto riesgo", mientras que las personas blancas que sí reincidieron eran clasificadas con mayor frecuencia como de bajo riesgo.

Los autores sostienen que el caso State v. Loomis, en el que se admitió la utilización de un algoritmo de código cerrado como fundamento para determinar una sentencia, evidencia una problemática esencial respecto del principio de igualdad ante la ley: al estar protegido el funcionamiento del sistema por el secreto comercial de su desarrollador, la defensa no pudo conocer

ni examinar la lógica interna del algoritmo, lo que restringió el ejercicio efectivo del derecho de contradicción.

Roa Avella et al. (2022) concluyen que la ausencia de transparencia algorítmica en materia penal no solo afecta la igualdad y la prohibición de discriminación, sino que también puede vulnerar garantías procesales fundamentales, entre ellas el debido proceso y el derecho a la defensa, y proponen prohibir o, al menos, someter a estrictos límites el empleo de algoritmos de código cerrado en decisiones judiciales susceptibles de afectar la libertad personal.

3.4. Discriminación algorítmica en el ámbito sanitario

En el ámbito sanitario, los estudios recientes alertan sobre el riesgo de que los algoritmos de apoyo a la decisión clínica produzcan sesgos estructurales. Amaya-Santos et al. (2024), desde un enfoque interseccional, muestran que cuando estos sistemas se entrenan con historiales médicos que subrepresentan a ciertos grupos —mujeres, minorías étnicas y personas de escasos recursos— tienden a perpetuar desigualdades ya existentes en el acceso y la calidad de la atención, pudiendo generar diagnósticos menos precisos o recomendaciones terapéuticas menos adecuadas para esas poblaciones.

Calahorrano Latorre (2025), desde una perspectiva jurídico-dogmática, complementa este análisis mediante la propuesta de mecanismos destinados a prevenir prácticas discriminatorias en la adopción de decisiones clínicas apoyadas en inteligencia artificial, entre los que se encuentran la validación externa y obligatoria de los sistemas antes de su implementación en entornos clínicos, la posibilidad de rastrear las variables utilizadas para generar una recomendación algorítmica y el reconocimiento del derecho del paciente a solicitar que la decisión sea revisada por un profesional humano. El autor señala que la confianza del paciente y la confiabilidad del sistema se debilitan cuando la herramienta funciona como una "caja negra", es decir, cuando su funcionamiento interno no puede ser comprendido ni explicado al personal sanitario ni al paciente.

Desde un enfoque filosófico, Morales Ramírez (2023) sitúa el sesgo algorítmico dentro de una antropología del pensamiento automático, señalando que la supuesta neutralidad de los algoritmos encubre elecciones de diseño con carga valorativa —como la selección de variables, la definición de qué cuenta como resultado "exitoso" o el umbral de error tolerable— que, en el fondo, reproducen la cosmovisión de quienes diseñan e implementan estos sistemas. Según la autora, la equidad algorítmica no puede entenderse como un mero ajuste técnico-estadístico, sino que exige una deliberación ética explícita sobre los valores que deben guiar el diseño de sistemas que inciden en derechos fundamentales.

3.5. El marco regulatorio europeo como respuesta normativa

El Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, denominado AI Act, representa uno de los avances regulatorios más significativos señalados por la literatura contemporánea. Martínez Navarro (2025) explica que la normativa clasifica los sistemas de inteligencia artificial en función del nivel de riesgo que representan, prohibiendo de manera expresa

las prácticas de puntuación social (social scoring) y el uso de sistemas que se aprovechen de las condiciones de vulnerabilidad de determinados grupos, especialmente cuando estas se relacionan con la edad, la discapacidad o la situación socioeconómica.

Asimismo, el reglamento exige identificar, valorar y prevenir los posibles daños que los sistemas de IA de alto riesgo puedan ocasionar a las personas o colectivos en situación de vulnerabilidad, incluyendo expresamente la accesibilidad para las personas con discapacidad y la promoción de la igualdad de género dentro de los aspectos que deben considerarse en dicha evaluación.

Capellà i Ricart (2024) valora el Reglamento (UE) 2024/1689 de forma favorable por reconocer, aunque de manera aún incipiente, el carácter estructural de la discriminación algorítmica, al imponer evaluaciones de impacto sobre derechos fundamentales para los sistemas de alto riesgo. No obstante, advierte que el texto sigue anclado en una lógica predominantemente anticlasificatoria y de gestión de riesgos individuales, sin integrar plenamente enfoques antisubordinatorios que aborden la dimensión colectiva y estructural del problema, como la concentración de poder informacional en pocos proveedores tecnológicos o la retroalimentación entre sistemas de distintos sectores, que puede amplificar a escala social patrones de exclusión que en cada aplicación individual parecen menores.

3.6. Experiencias regulatorias emergentes fuera del contexto europeo

Conviene situar el Reglamento (UE) 2024/1689 dentro de un panorama regulatorio más amplio y todavía en formación. En Brasil, el Proyecto de Ley 2338/2023, aprobado por el Senado Federal el 10 de diciembre de 2024 y remitido a la Cámara de Diputados, donde a la fecha de cierre de este capítulo continúa en tramitación en una comisión especial (Senado Federal de Brasil, 2023), propone, a semejanza del modelo europeo, una clasificación de los sistemas de IA en función del riesgo —que distingue los usos de riesgo excesivo, prohibidos, de los de alto riesgo, sujetos a obligaciones reforzadas—, la creación de un Sistema Nacional de Regulación y Gobernanza de la Inteligencia Artificial coordinado por la Autoridad Nacional de Protección de Datos y el reconocimiento expreso de un catálogo de derechos de las personas afectadas, entre ellos los derechos a la explicación, a la impugnación y a la revisión humana de las decisiones automatizadas, así como a la no discriminación ilícita o abusiva y a la corrección de sesgos discriminatorios.

A diferencia del AI Act, el proyecto brasileño incorpora, además, un régimen específico de responsabilidad civil y reglas sobre la remuneración por el uso de contenidos protegidos en el entrenamiento de los modelos, lo que refleja una síntesis particular entre la lógica europea de gestión de riesgos y preocupaciones propias del contexto latinoamericano.

En los Estados Unidos, ante la ausencia de una legislación federal integral, la respuesta regulatoria se ha desplazado hacia el nivel estatal, siendo Colorado el caso más representativo. La Ley de Inteligencia Artificial de Colorado (SB 24-205), sancionada en 2024, imponía a los desarrolladores y a los implementadores de sistemas de IA de alto riesgo un deber de diligencia razonable para proteger a las personas consumidoras frente a la discriminación algorítmica en

decisiones consecuentes relativas al empleo, la vivienda, la salud, el crédito o los servicios legales (Estado de Colorado, 2024).

Este marco fue objeto de sucesivas postergaciones y, finalmente, derogado y reemplazado en mayo de 2026 por la Ley de Tecnología de Decisión Automatizada (SB 26-189), aplicable a partir del 1 de enero de 2027, de alcance más acotado y centrada en obligaciones de transparencia, aviso previo al consumidor, explicación tras una decisión adversa, corrección de datos y revisión humana, y que atribuye la responsabilidad por discriminación conforme al derecho antidiscriminatorio estatal general (Estado de Colorado, 2026).

La trayectoria regulatoria de Colorado ilustra una tensión que la literatura revisada también documenta para el contexto europeo: la dificultad de sostener, en el tiempo y frente a la presión de la industria, obligaciones robustas de gestión de riesgos y auditoría orientadas específicamente a la no discriminación algorítmica, frente a modelos más limitados centrados en la transparencia y la notificación.

Estas experiencias, aunque todavía incipientes y sujetas a modificaciones, confirman que el problema de la discriminación algorítmica no es privativo del contexto europeo y que existen, al menos, dos vías regulatorias alternativas: una de corte integral y preventivo, inspirada en la gestión de riesgos por niveles (Unión Europea, Brasil), y otra de corte más sectorial y reactivo, centrada en deberes de cuidado, transparencia y remedios individuales (Estados Unidos). Para América Latina, la experiencia brasileña constituye un referente comparado directo que, sin embargo, contrasta con la práctica ausencia de proyectos normativos equivalentes en la mayoría de los países de la región.

Más allá de esta clasificación general, las diferencias sustantivas entre los tres modelos tienen implicaciones directas para la protección del principio de igualdad (Tabla 5). En primer lugar, difieren en el *fundamento* de la protección: el Reglamento (UE) 2024/1689 se estructura como una norma de seguridad de producto basada en la gestión de riesgos, en la que la no discriminación figura entre los derechos fundamentales que deben protegerse, pero no como su eje organizador; el proyecto brasileño, en cambio, enuncia de forma expresa un catálogo de derechos de las personas afectadas que incluye la no discriminación y la corrección de sesgos; y la ley vigente en Colorado se limita a garantizar transparencia y vías de revisión, remitiendo la tutela antidiscriminatoria sustantiva a la legislación general.

En segundo lugar, difieren en el *momento* de la intervención: los modelos europeo y brasileño son preventivos, pues exigen evaluar y mitigar riesgos antes del despliegue, mientras que el modelo de Colorado es predominantemente reactivo y se activa tras una decisión adversa.

En tercer lugar, difieren en la *distribución de la carga*: en los dos primeros, la obligación de detectar y mitigar sesgos recae principalmente en proveedores e implementadores, mientras que en el tercero la iniciativa recae, en buena medida, en la persona afectada, que debe solicitar explicaciones, corrección o revisión humana. Desde la perspectiva antisubordinatoria descrita por Capellà i Ricart (2024), esta diferencia es decisiva: los modelos reactivos e individualizados son estructuralmente menos aptos para captar la discriminación sistémica y acumulativa, porque

dependen de que las personas afectadas, con frecuencia, las que menos recursos tienen para ello, identifiquen el daño y lo reclamen.

Tabla 5

Comparación de los modelos regulatorios analizados desde la perspectiva del principio de igualdad

Dimensión	Unión Europea: Reglamento (UE) 2024/1689	Brasil: Proyecto de Ley 2338/2023	Colorado: SB 24-205 y SB 26-189
Estado normativo	En vigor desde el 1 de agosto de 2024, con aplicación escalonada	Aprobado por el Senado (2024); en tramitación en la Cámara de Diputados	SB 24-205 derogada antes de su aplicación; SB 26-189 aplicable desde el 1 de enero de 2027
Lógica regulatoria	Gestión de riesgos por niveles, con enfoque de seguridad de producto	Gestión de riesgos por niveles combinada con un catálogo de derechos	Transparencia y garantías procedimentales frente a decisiones automatizadas
Momento de intervención	Preventivo (ex ante)	Preventivo (ex ante) y reparador	Predominantemente reactivo (ex post)
Tratamiento de la no discriminación	Derecho fundamental protegido mediante gestión de riesgos, gobernanza de datos y evaluación de impacto	Derecho expreso a la no discriminación ilícita o abusiva y a la corrección de sesgos	Remitida al derecho antidiscriminatorio estatal general
Derechos de las personas afectadas	Explicación de decisiones individuales adoptadas con sistemas de alto riesgo y reclamación ante la autoridad	Explicación, impugnación y revisión humana	Aviso previo, explicación tras decisión adversa, corrección de datos y revisión humana
Supervisión	Autoridades nacionales de vigilancia y Oficina Europea de IA	Sistema Nacional de Regulación y Gobernanza de la IA, coordinado por la autoridad de protección de datos	Fiscalía General del estado
Implicación para la igualdad	Protección preventiva, pero centrada en riesgos individuales; integración incompleta de la dimensión estructural	Mayor protección sustantiva, condicionada a su aprobación definitiva	Protección dependiente de la iniciativa individual; débil frente a la discriminación estructural

Nota. El estado normativo corresponde a la fecha de cierre del capítulo (octubre de 2026). Elaboración propia a partir de Parlamento Europeo y Consejo de la Unión Europea (2024), Senado Federal de Brasil (2023), Estado de Colorado (2024, 2026) y Martínez Navarro (2025).

El análisis del corpus documental revela vacíos de conocimiento que deben orientar la agenda investigativa futura. En primer lugar, se evidencia una marcada asimetría geográfica: la mayor parte de los estudios se concentra en la Unión Europea, el Reino Unido y Estados Unidos, mientras que la producción académica sobre discriminación algorítmica en América Latina es comparativamente reducida, pese a que diversos países de la región ya implementan sistemas de inteligencia artificial en ámbitos sensibles como la asignación de subsidios sociales o la vigilancia policial predictiva; solo en Colombia, por ejemplo, se han documentado más de un centenar de sistemas de decisión automatizada en entidades públicas (Gutiérrez & Castellanos, 2023).

Salvo la excepción parcial del proyecto de ley brasileño, la región carece de marcos normativos específicos en materia de discriminación algorítmica comparables al Reglamento (UE) 2024/1689, lo que se traduce en un doble déficit: escasa evidencia empírica sobre el funcionamiento efectivo de los sistemas desplegados en la región y ausencia de estándares jurídicos propios frente a los cuales evaluar dicho funcionamiento.

Esta concentración limita la capacidad de la literatura para reflejar las particularidades socioeconómicas, institucionales y de infraestructura de datos de contextos con menor capacidad regulatoria, y plantea el riesgo de que las soluciones técnicas y normativas diseñadas para contextos europeos o estadounidenses se trasladen de forma acrítica a realidades institucionales sustancialmente distintas.

En segundo término, subsiste una fragmentación metodológica entre los enfoques jurídico-doctrinales, que analizan el fenómeno desde la interpretación normativa sin incorporar evidencia empírica sobre el funcionamiento efectivo de los sistemas, y los estudios técnicos de equidad algorítmica, que suelen reducir la noción de discriminación a la paridad estadística entre grupos, sin profundizar en las categorías jurídicas de discriminación directa, indirecta o por asociación. La literatura revisada apunta a la necesidad de fortalecer el diálogo metodológico entre ambas tradiciones, de modo que los estándares jurídicos de igualdad se traduzcan en métricas técnicas auditables y, a la inversa, los hallazgos de las auditorías algorítmicas adquieran sentido jurídico.

En tercer lugar, se observa una laguna en torno a la discriminación algorítmica interseccional. Aunque investigaciones recientes, como la de Amaya-Santos et al. (2024), han comenzado a examinar de qué manera la combinación de múltiples categorías de vulnerabilidad agrava el riesgo de discriminación, la mayoría de las métricas técnicas de equidad siguen concebidas para evaluar una única variable protegida de forma aislada. Esto obstaculiza la identificación de patrones de discriminación compuesta, los cuales únicamente se hacen evidentes al analizar subgrupos definidos por la intersección de dos o más características.

En cuarto lugar, la literatura revela una atención desproporcionada entre los colectivos estudiados: mientras el género, la discapacidad y la raza cuentan con un tratamiento consolidado, las personas mayores y la infancia y adolescencia permanecen en un plano secundario. Los escasos trabajos disponibles sobre estos grupos tienden a enfocarse en riesgos colaterales como la privacidad, la manipulación comercial o el edadismo en sistemas de salud, sin articular de manera plena el análisis bajo el principio de igualdad y no discriminación. Esta asimetría se reproduce en el propio corpus analizado: ninguno de los estudios incluidos tiene como objeto principal a las personas mayores o a niñas, niños y adolescentes, y los colectivos migrante y afrodescendiente cuentan con un solo estudio cada uno (Iturmendi Rubia, 2023; Roa Avella et al., 2022).

Finalmente, se identifica un vacío de gobernanza relativo a los mecanismos efectivos de auditoría algorítmica externa, independiente y periódica. Aunque el Reglamento (UE) 2024/1689 establece obligaciones de evaluación de conformidad para los sistemas de alto riesgo, y el proyecto brasileño incorpora un organismo regulador con funciones análogas, la doctrina subraya que su aplicación real dependerá de la capacidad institucional de los organismos de supervisión, aún en

proceso de configuración, así como de la disponibilidad de metodologías de auditoría estandarizadas y del acceso efectivo de investigadores independientes a los sistemas evaluados.

Hasta ahora, estos avances han enfrentado barreras derivadas del secreto comercial y de la protección de la propiedad intelectual de los desarrolladores, tal como ilustra el caso COMPAS documentado por Roa Avella et al. (2022). La literatura coincide en que, sin acceso técnico efectivo a los sistemas, más allá de la mera obligación formal de someterlos a evaluación, la auditoría algorítmica corre el riesgo de convertirse en un requisito documental antes que, en una garantía sustantiva de no discriminación, lo que exige a los organismos de supervisión desarrollar capacidades técnicas propias, y no depender exclusivamente de la autoevaluación reportada por los proveedores de los sistemas.

4. Discusión

La principal aportación de esta revisión no reside en identificar casos nuevos de discriminación algorítmica, sino en integrar, bajo el prisma del principio de igualdad, una evidencia hasta ahora fragmentada por disciplinas y sectores, y en ofrecer un esquema (Tabla 4) que permite distinguir analíticamente sus mecanismos de producción y sus formas jurídicas. Para interpretar correctamente este aporte es necesario diferenciar dos niveles en los hallazgos.

El primero corresponde a lo *documentado empíricamente*: la existencia de disparidades verificadas en sistemas concretos, como la clasificación diferencial del riesgo de reincidencia en COMPAS (Roa Avella et al., 2022) o la penalización de currículos asociados a mujeres en un sistema de selección de personal (Lousada Arochena, 2024), y la identificación de mecanismos de sesgo concretos, de muestreo, de reflejo de prejuicios y de interacción (Pérez, 2024).

El segundo corresponde a *construcciones interpretativas* derivadas de la síntesis teórica: la caracterización de la discriminación algorítmica como fenómeno estructural, la tesis de la "monocultura algorítmica" (Capellà i Ricart, 2024) o la idea de que los sistemas de IA tienden a amplificar, más que a corregir, las desigualdades preexistentes. Estas últimas proposiciones cuentan con respaldo convergente en la literatura revisada, pero descansan mayoritariamente en argumentación doctrinal y en la extrapolación a partir de un número limitado de casos paradigmáticos, por lo que deben entenderse como hipótesis interpretativas plausibles y no como generalizaciones empíricamente demostradas.

La comparación entre autores revela, además, divergencias que no deben minimizarse. La primera enfrenta a quienes confían en la corrección técnica del sesgo con quienes la consideran insuficiente. Calahorrano Latorre (2025) y El Morr et al. (2024) privilegian soluciones centradas en la representatividad de los datos, la validación externa de los sistemas y el diseño de tecnologías accesibles; Tilmes (2022), en cambio, advierte que la paridad estadística entre grupos resulta inadecuada para un colectivo tan heterogéneo como el de las personas con discapacidad y propone complementarla con ajustes razonables individualizados; y Morales Ramírez (2023) y Capellà i Ricart (2024) sostienen que ninguna calibración técnica es suficiente sin una deliberación explícita

sobre los valores incorporados en el diseño y sin alterar las estructuras de poder que determinan quién diseña los sistemas, con qué finalidad y bajo qué modelo de negocio.

La elección no es neutral en términos de política pública: privilegiar exclusivamente la auditoría técnica traslada la carga hacia estándares de ingeniería que los organismos de supervisión de la región todavía no han desarrollado, mientras que privilegiar exclusivamente la limitación normativa de usos puede resultar insuficiente frente a sistemas ya desplegados cuyo funcionamiento interno permanece opaco.

La segunda divergencia se refiere a la capacidad del derecho antidiscriminatorio vigente para abordar el fenómeno. Adams-Prassl et al. (2023) sostienen que buena parte de los casos puede reconducirse a la categoría tradicional de discriminación directa cuando el algoritmo opera como un sustituto prácticamente perfecto de una característica protegida. Lousada Arochena (2024) se pregunta, por el contrario, si es necesario un concepto autónomo de discriminación algorítmica, y Capellà i Ricart (2024) va más allá al sostener que el enfoque anticlasificatorio, centrado en casos individuales, es estructuralmente incapaz de captar los efectos acumulativos de varios sistemas aparentemente neutros sobre un mismo grupo.

El contraste sugiere que el derecho vigente es relativamente eficaz frente a la discriminación directa y frente a la indirecta claramente identificable, pero débil frente a la discriminación sistémica, que es precisamente la que la literatura considera más característica de los entornos algorítmicos.

Estas divergencias se proyectan sobre el debate regulatorio. La literatura revisada coincide en que una obligación normativa de evaluación, como la prevista en el Reglamento (UE) 2024/1689 o en el proyecto de ley brasileño, constituye una condición necesaria, pero no suficiente: el caso COMPAS muestra que el secreto comercial puede neutralizar el derecho de defensa incluso en jurisdicciones con reglas procesales garantistas (Roa Avella et al., 2022), y la trayectoria de Colorado ilustra cómo los mecanismos más exigentes de gestión de riesgos tienden a debilitarse ante la presión de la industria cuando carecen de respaldo institucional.

La eficacia de la auditoría algorítmica depende, por tanto, menos de la existencia formal de la obligación que de su diseño concreto: acceso técnico efectivo de auditores independientes a los sistemas, periodicidad de la revisión, consecuencias jurídicas claras ante el incumplimiento y canales accesibles para que los propios colectivos afectados puedan activarla.

La escasez de evidencia latinoamericana tiene, a su vez, implicaciones metodológicas que condicionan la transferibilidad de estos hallazgos a la región. En primer lugar, los mecanismos de producción del sesgo dependen del contexto de los datos: en países con altos niveles de informalidad laboral, subregistro estadístico de poblaciones indígenas y afrodescendientes o brechas de conectividad, los sesgos de muestreo y de medición pueden adoptar formas y magnitudes distintas de las documentadas en Europa o Estados Unidos, de modo que la ausencia de estudios regionales no equivale a la ausencia del problema, sino al desconocimiento de su configuración específica.

En segundo lugar, los remedios propuestos por la literatura, auditorías independientes, evaluaciones de impacto, autoridades de supervisión especializadas, presuponen una capacidad

institucional que no puede darse por supuesta; la evidencia colombiana muestra, por ejemplo, que aun existiendo normas que exigen transparencia algorítmica, los repositorios públicos de algoritmos resultan incompletos e inconsistentes y el acceso al código de los sistemas se ha denegado por razones de privacidad, seguridad o propiedad intelectual (Gutiérrez & Castellanos, 2023).

En tercer lugar, el reducido número de estudios latinoamericanos del corpus, dos de los cuales analizan marcos normativos o casos extranjeros, impide distinguir qué hallazgos son generalizables y cuáles son propios de los contextos de origen; trabajos regionales sobre racismo algorítmico, como el de Silva (2023) en Brasil, sugieren además que los ejes de raza y clase pueden tener en la región un peso relativo distinto del que les asigna la literatura europea. Por ello, respecto de América Latina, las conclusiones de esta revisión deben leerse como hipótesis de trabajo que requieren contraste empírico local, y no como diagnósticos directamente aplicables.

Finalmente, la interpretación de estos resultados debe considerar tres limitaciones del conocimiento disponible. La primera es la dependencia de la literatura doctrinal: la mayor parte del corpus está compuesta por estudios jurídico-doctrinales, teóricos o filosóficos, y ninguno de los trabajos incluidos consiste en una auditoría directa de sistemas desplegados en contextos reales. Ello implica que muchas afirmaciones sobre el funcionamiento discriminatorio de los sistemas descansan en un reducido número de casos paradigmáticos —COMPAS o el cribado curricular sesgado por género— que se reiteran entre autores, con el consiguiente riesgo de circularidad argumentativa, y que la literatura jurídica tiende a razonar sobre riesgos potenciales más que sobre efectos medidos.

Los estudios empíricos recientes ajenos al corpus, como las auditorías de modelos de lenguaje en procesos de selección (An et al., 2025; Wilson & Caliskan, 2024) o los análisis de modelos de diagnóstico por imagen que muestran que la equidad alcanzada en un entorno no se mantiene necesariamente en otros hospitales y poblaciones (Yang et al., 2024), confirman la existencia de disparidades, pero también evidencian que su dirección y su magnitud varían según el sistema, el contexto y el subgrupo, lo que aconseja cautela frente a las generalizaciones.

La segunda limitación es el desarrollo desigual de los colectivos analizados: el género, la discapacidad y la justicia penal concentran la mayor parte de la evidencia, mientras que las personas mayores y la infancia y adolescencia, aunque mencionadas como colectivos vulnerables, carecen de estudios específicos en el corpus, y las personas migrantes y afrodescendientes cuentan con un solo estudio cada una; los hallazgos, por tanto, no pueden extenderse sin más a estos grupos. La tercera es la escasez de estudios longitudinales que evalúen si las medidas de mitigación reducen efectivamente la discriminación a lo largo del tiempo, unida a la persistente dificultad para traducir el concepto jurídico de discriminación indirecta en métricas computacionalmente auditables.

5. Conclusiones

1. La evidencia revisada permite afirmar que los sistemas de inteligencia artificial pueden reproducir y, en determinadas condiciones, amplificar desigualdades presentes en los datos de entrenamiento, con efectos desproporcionados documentados sobre mujeres, personas

con discapacidad y personas afrodescendientes en los ámbitos del empleo, la salud y la justicia penal. La caracterización de la discriminación algorítmica como un riesgo estructural, y no meramente incidental, para el principio de igualdad constituye, en cambio, una conclusión interpretativa de la síntesis teórica, ampliamente compartida por la literatura doctrinal, pero que requiere mayor contraste empírico en sistemas desplegados en contextos reales.

2. La revisión evidencia que la discriminación algorítmica adopta manifestaciones sectoriales específicas en el empleo, la salud, la justicia penal y la protección migratoria, cuyos mecanismos de producción y cuyas consecuencias jurídicas no pueden analizarse de manera uniforme, lo que exige marcos de intervención regulatoria y técnica sensibles a las particularidades de cada ámbito de aplicación.

3. La distinción entre sesgo algorítmico, inequidad estadística, discriminación algorítmica y vulneración del principio de igualdad permite situar con precisión cada hallazgo, y los enfoques teóricos anticlasificatorio, antisubordinatorio, interseccional y de justicia distributiva ofrecen lentes complementarias, más que excluyentes, para comprender el fenómeno, según se sintetiza en la Tabla 4; su integración resulta necesaria para abordar tanto las manifestaciones individuales de discriminación directa como sus efectos estructurales y acumulativos sobre grupos históricamente subordinados, incluida la discriminación compuesta, que las métricas de equidad centradas en una sola variable no logran detectar.

4. El análisis comparado muestra dos vías regulatorias con implicaciones distintas para la igualdad: una preventiva, basada en la gestión de riesgos por niveles (Unión Europea y Brasil), y otra predominantemente reactiva, centrada en la transparencia y en remedios individuales (Colorado). La primera ofrece una protección más robusta frente a la discriminación estructural, especialmente cuando, como en el proyecto brasileño, se acompaña de derechos expresos de no discriminación; sin embargo, en ambos casos su eficacia real dependerá de la capacidad institucional para desarrollar mecanismos de auditoría algorítmica externa, independiente y con acceso efectivo a los sistemas evaluados, un desafío de gobernanza todavía pendiente de resolución.

5. La contribución original de esta revisión consiste en sistematizar, bajo un marco común centrado en el principio de igualdad, evidencia dispersa entre disciplinas y sectores de aplicación; en ofrecer, a través de la Tabla 4, un esquema integrado que relaciona explícitamente los enfoques teóricos, los mecanismos de producción del sesgo y las manifestaciones sectoriales de la discriminación algorítmica; y en examinar comparativamente los modelos regulatorios europeo, brasileño y de Colorado, así como las condiciones de transferibilidad de la evidencia a América Latina.

6. Entre las limitaciones más relevantes del conocimiento existente destacan el predominio de estudios doctrinales sobre estudios empíricos de sistemas desplegados, la concentración geográfica de la evidencia en contextos europeos y estadounidenses, la escasez de estudios sobre discriminación algorítmica interseccional y el desarrollo desigual de los colectivos

analizados: las personas mayores y la infancia y adolescencia carecen de estudios específicos en el corpus, y las personas migrantes y afrodescendientes cuentan con evidencia muy limitada, por lo que las conclusiones anteriores no pueden extrapolarse a estos grupos sin investigación adicional.

5.1. Recomendaciones e implicaciones normativas

Las recomendaciones que siguen no se derivan directamente de la evidencia revisada, sino que constituyen propuestas de carácter normativo elaboradas a partir de la síntesis realizada. Se presentan separadas de las conclusiones para distinguir con claridad lo que la literatura permite afirmar de lo que, a juicio de quien escribe, debería hacerse.

1. Se recomienda a los órganos legislativos de la región que, al diseñar marcos normativos sobre inteligencia artificial, incorporen de manera expresa un deber de no discriminación algorítmica exigible frente a sistemas de alto riesgo desplegados en sectores sensibles (salud, empleo, justicia penal, programas sociales), tomando como referentes comparados tanto el Reglamento (UE) 2024/1689 como el Proyecto de Ley 2338/2023 de Brasil y la evolución reciente de la legislación estatal de Colorado, evaluando críticamente las fortalezas y debilidades de cada modelo antes de trasplantarlo.
2. Se recomienda a los organismos de control y supervisión que desarrollen capacidades técnicas propias para la auditoría algorítmica independiente, que no dependan exclusivamente de la autoevaluación reportada por los proveedores de los sistemas, y que habiliten canales accesibles para que los colectivos vulnerables y sus representantes puedan solicitar la revisión de decisiones automatizadas que los afecten, incorporando expresamente el principio de ajustes razonables para las personas con discapacidad.
3. Se recomienda a las personas y equipos desarrolladores de sistemas de inteligencia artificial adoptar prácticas de equidad desde el diseño (fairness by design), documentar de manera trazable el origen y la representatividad de los datos de entrenamiento, incorporar pruebas de equidad interseccional, y no únicamente evaluaciones de paridad estadística sobre una sola variable protegida, y habilitar mecanismos efectivos de revisión humana ante decisiones consecuentes para las personas afectadas.
4. Se recomienda que futuras líneas de investigación profundicen en estudios empíricos y comparados en contextos latinoamericanos, incluyendo el seguimiento de la tramitación del Proyecto de Ley 2338/2023 en Brasil y de eventuales iniciativas similares en otros países de la región, en el desarrollo de métricas de equidad algorítmica capaces de capturar la discriminación interseccional, y en la construcción de metodologías de auditoría que traduzcan de manera operativa las categorías jurídicas de discriminación directa e indirecta al lenguaje técnico de la ingeniería de sistemas, favoreciendo así un diálogo interdisciplinario más robusto entre el derecho, la ética y la ciencia de datos.

Referencias bibliográficas

Adams-Prassl, J., Binns, R., & Kelly-Lyth, A. (2023). Algoritmos directamente discriminatorios. *The Modern Law Review, 86*(1), 144-175. https://doi.org/10.1111/1468-2230.12759

Álvarez García, H. (2023). La discriminación algorítmica por razón de discapacidad. *Revista Internacional de Pensamiento Político*(18), 15-39. https://doi.org/10.46661/revintpensampolit.8103

Amaya-Santos, S., Jiménez-Pernett, J., & Bermúdez-Tamayo, C. (2024). ¿Salud para quién? Interseccionalidad y sesgos de la inteligencia artificial para el diagnóstico clínico. *Anales del Sistema Sanitario de Navarra, 47(2), e1077. https://doi.org/10.23938/ASSN.1077*

An, J., Huang, D., Chen, L., & Tai, M. (2025). Medición de sesgos de género y raciales en modelos lingüísticos a gran escala: evidencia interseccional a partir de la evaluación automatizada de currículos. *PNAS Nexus, 4*(3), 1-14. https://doi.org/10.1093/pnasnexus/pgaf089

Braun, V., & Clarke, V. (2021). Thematic analysis: A practical guide. SAGE.

Calahorrano Latorre, E. (2025). Equidad algorítmica y decisiones clínicas basadas en sistemas de soporte basados en inteligencia artificial. *Nuevo Derecho, 21*(36), 1-17. https://doi.org/10.25057/2500672X.1691

Capellà i Ricart, A. (2024). Antisubordinación y discriminación algorítmica. *Revista Pensamiento Jurídico*(37), 214-237. https://doi.org/10.36151/TD.2024.112

El Morr, C., Kundi, B., Sara Taleghani, F. M., El-Lahib, Y., & Gorman, R. (2024). Inteligencia artificial y discapacidad: una revisión sistemática exploratoria. *Sage Journals*(1), 1-24. https://doi.org/10.1177/1460458224128574

Estado de Colorado. (2024, 17 de mayo). Senate Bill 24-205. https://leg.colorado.gov/bills/sb24-205

Estado de Colorado. (2026, 14 de mayo). Senate Bill 26-189. https://leg.colorado.gov/bills/sb26-189

Flores Anarte, L. (2023). Sesgos de género en la inteligencia artificial: El estado de derecho frente a la discriminación algorítmica por razón de sexo. *Revista Internacional de Pensamiento Político*(18), 97-122. https://doi.org/10.46661/revintpensampolit.8778

Foulds, J. R., Islam, R., Naher Keya, K., & Pam, S. (2020). Una definición interseccional de equidad. *2020 IEEE 36th International Conference on Data Engineering (ICDE).* Dallas: IEEE. https://doi.org/10.1109/ICDE48307.2020.00203

Gutiérrez, J. D., & Castellanos Sánchez, M. (2023). Transparencia algorítmica y estado abierto en Colombia. *Reflexión Política, 25(52), 6-21*. https://doi.org/10.29375/01240781.4789

Iturmendi Rubia, J. M. (2023). La discriminación algorítmica y su impacto en la dignidad de la persona y los derechos humanos. Especial referencia a los migrantes. *Deusto Derechos Humanos*(11), 257-284. https://doi.org/10.18543/djhr.2910

Kearns, M., Neel, S., Roth, A., & Steven Wu, Z. (2018). Preventing Fairness Gerrymandering: Auditing and learning for subgroup fairness. *Proceedings on the 35th.* Stockholm: PMLR 80. https://proceedings.mlr.press/v80/kearns18a.html

Lousada Arochena, J. F. (2024). Inteligencia artificial y sesgos discriminatorios: ¿es necesario un nuevo concepto de discriminación algorítmica? *IgualdaDES*(11), 97-123. https://doi.org/10.18042/cepc/IgdES.11.04

Martínez Navarro, J. A. (2025). Inteligencia artificial, discriminación algorítmica y la respuesta normativa de la Unión Europea. *Ces Derecho, 16*(3), 4-20. https://doi.org/10.21615/cesder.7943

Morales Ramírez, G. (2023). Problemática antropológica detrás de la discriminación generada a partir de los algoritmos de inteligencia artificial. *Medicina y Ética, 2*(34), 429-480. https://doi.org/10.36105/mye.2023v34n2.04

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*(71), 1-9. https://doi.org/10.1136/bmj.n71

Parlamento Europeo y Consejo de la Unión Europea. (2024, 13 de junio). *Reglamento (UE) 2024/1689.* https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Pérez-Ugena Coromina, M. (2024). Sesgo de género en IA. *Revista en cultura de la legalidad*(26), 311-330. https://doi.org/10.20318/eunomia.2024.8515

Roa Avella, M. P., Sanabria Moyano, J., & Dinas Hurtado, K. (2022). Uso de algoritmos COMPAS en el proceso penal y los riesgos a los derechos humanos. *Revista Brasileira de Direito Processual Penal, 8*(1), 275-310. https://doi.org/10.22197/rbdpp.v8i1.615

Senado Federal de Brasil. (2023, 11 de agosto). *Proyecto de Ley 2338/2023.* https://www25.senado.leg.br/web/atividade/materias/-/materia/157233

Silva, T. (2023). Racismo algorítmico: Inteligencia artificial y discriminación en redes sociales. *Em favor de Igualdade racial, 6*(3), 188-192. https://doi.org/10.29327/269579.6.3-14

Tilmes, N. (2022). Discapacidad, equidad, y sesgo algorítmico en el reclutamiento mediante IA. *Ethics and Information Technology, 24*(21). https://doi.org/10.1007/s10676-022-09633-2

Wilson, K., & Caliskan, A. (2024). Gender, race, and intersectional bias in resume screening via language model retrieval. *Ética y Sociedad, 7*(1), 1578-1590. https://doi.org/10.1609/aies.v7i1.31748

Yang, Y., Zhang, H., Gichoya, J. W., Katabi, D., & Ghassemi, M. (2024). The limits of fair medical imaging AI in real-world generalization. *Nature Medicine, 30*(10), 2838–2848. https://doi.org/10.1038/s41591-024-03113-4