

Dossier

## Algoritmos Discriminadores y Mercados de Trabajo Ineficientes: Sesgos y discriminación en la contratación y evaluación de desempeño laboral. La Contratación bajo el paradigma Post- Homo economicus

Discriminatory Algorithms and Inefficient Labor Markets: Biases and Discrimination in Hiring and Performance Evaluation. Hiring under the Post-Homo Economicus Paradigm

Algoritmos Discriminatórios e Mercados de Trabalho Ineficientes: Vieses e Discriminação na Contratação e Avaliação de Desempenho. Contratação sob o Paradigma Pós-Homo Economicus



Carolina Pinasco  
Universidad Nacional de Río Cuarto, Argentina  
cpinasco@gmail.com

**Resumen** Este artículo explica cómo herramientas algorítmicas en selección y evaluación laboral pueden generar impacto dispar y fallos de eficiencia en mercados de trabajo. Integra análisis económico del derecho (discriminación por preferencias y estadística, costos de transacción, asimetrías de información) con la dimensión jurídico-procesal de casos recientes y marcos regulatorios. Aporta una arquitectura de gobernanza con auditoría ex-ante y trazabilidad, y explicita un plan de evaluación econométrica replicable para medir efectos en tasas de selección y brechas salariales por grupo. Sustenta que la mitigación de sesgos es condición de eficiencia, no solo de cumplimiento.

**Palabras clave:** discriminación algorítmica, análisis económico del derecho, impacto dispar, asimetría de información, costos de transacción, gobernanza algorítmica.

**Abstract** The paper shows how algorithmic tools in hiring and performance evaluation can produce disparate impact and efficiency losses. It integrates law-and-economics with case-law and comparative regulation. It proposes a governance architecture combining ex-ante audits and traceability, and provides a replicable econometric plan to estimate group-level disparities in selection rates and wages. Bias mitigation is reframed as an efficiency requirement, not merely compliance.

**Keywords:** algorithmic discrimination, law & economics, disparate impact, information asymmetry, transaction costs, algorithmic governance.

**Resumo** Este artigo explica como ferramentas algorítmicas na seleção e avaliação de empregos podem gerar impactos díspares e falhas de eficiência nos mercados de trabalho. Integra a análise econômica do direito (discriminação com base em preferências e estatísticas, custos de transação, assimetrias de informação) com a dimensão jurídica e processual de casos recentes e marcos regulatórios. Apresenta uma arquite-

**DOI:** <https://doi.org/10.63207/a84d1744>

**Recepción:** 12 octubre 2025

**Aprobación:** 01 diciembre 2025



Revista Fundamentos

núm. 2, 2025

Universidad Nacional de Río Cuarto,  
Argentina

**ISSN-E:** 2545-6318

[fundamentos@fce.unrc.edu.ar](mailto:fundamentos@fce.unrc.edu.ar)

Carolina Pinasco

Universidad Nacional de Río  
Cuarto, Argentina  
[cpinasco@gmail.com](mailto:cpinasco@gmail.com)

Periodicidad: Semestral

URL: [https://  
portal.amelica.org/  
ameli/journal/  
966/9665479011/](https://portal.amelica.org/ameli/journal/)

[cpinasco@gmail.com](mailto:cpinasco@gmail.com)

tura de governança com auditoria ex-ante e rastreabilidade, e explica um plano de avaliação econométrica replicável para mensurar os efeitos nas taxas de seleção e nas disparidades salariais por grupo. Argumenta que a mitigação de vieses é uma condição para a eficiência, não apenas para a conformidade.

**Palavras-chave:** Discriminação algorítmica, análise econômica do direito , impacto desigual, assimetria de informação, custos de transação, governança algorítmica.

## Introducción

En la era de la transformación digital, los sistemas de inteligencia artificial (IA) se han convertido en una pieza central de la gestión empresarial, especialmente en áreas sensibles como la selección, evaluación y promoción de personas. En el discurso público, estas herramientas suelen presentarse como instrumentos neutrales orientados a aumentar la eficiencia, reducir costos y dejar atrás los prejuicios humanos. Sin embargo, la experiencia acumulada en la última década sugiere un escenario más complejo: lejos de eliminar las viejas formas de discriminación, ciertos sistemas algorítmicos tienden a reproducirlas y, en algunos casos, a amplificarlas a gran escala.

En el ámbito del empleo, la adopción de sistemas de IA se ha producido con mucha mayor rapidez que la construcción de marcos normativos, capacidades técnicas y mecanismos de rendición de cuentas adecuados. Empresas y organismos públicos incorporan herramientas de filtrado de currículums, entrevistas automatizadas o scoring de riesgo laboral sin comprender del todo sus supuestos de diseño, sus limitaciones y sus efectos distributivos. Esa brecha de comprensión genera, a la vez, oportunidades de innovación y riesgos significativos para la igualdad de oportunidades, la transparencia y la protección de datos personales.

Este artículo se inserta en ese contexto y parte de una pregunta central: ¿cómo es posible que herramientas diseñadas para mejorar la eficiencia y reducir la discrecionalidad humana terminen generando discriminación y mercados de trabajo más ineficientes? Para responderla, se integra el análisis económico de la discriminación con la literatura reciente sobre sesgos algorítmicos y con casos judiciales paradigmáticos. Sobre esa base, se argumenta que ciertos usos de la IA en recursos humanos configuran un fallo de mercado de carácter sistemático —la “discriminación algorítmica facilitada”— y se propone un modelo regulatorio dual que combine auditorías de sesgos con análisis de impacto económico y un régimen de responsabilidad reforzada para proveedores de sistemas de alto riesgo.

## 1. LA PARADOJA DE LA EFICIENCIA Y LA EMERGENCIA DE LA MACHINA ECONÓMICA

### 1.1. Contextualización: La “Discriminación Algorítmica Facilitada” como Problema Sistémico

Cuando pensamos en prejuicios y discriminación, solemos asociarlos con decisiones abiertamente hostiles o con estereotipos explícitos sobre raza, género, edad u otras características protegidas. Sin embargo, una parte importante de la discriminación cotidiana no adopta la forma de odio manifiesto, sino de atajos cognitivos, sesgos implícitos y reglas informales que se activan de manera automática cuando se decide a quién entrevistar, a quién promover o a quién despedir. Estos mecanismos son difíciles de detectar a nivel

individual, pero producen patrones agregados sistemáticos de exclusión.

La investigación en psicología cognitiva y economía comportamental ha mostrado que la mente humana recurre a heurísticas para simplificar decisiones complejas bajo incertidumbre. Esas heurísticas pueden ser funcionales en ciertos contextos, pero se vuelven problemáticas cuando se aplican a decisiones masivas de alto impacto, como la contratación o la asignación de oportunidades laborales. La discriminación deja entonces de ser un acto excepcional para convertirse en una característica recurrente de la toma de decisiones.

La transformación digital de los mercados de trabajo introduce una paradoja fundamental. Las organizaciones recurren a sistemas algorítmicos con la expectativa de reducir la arbitrariedad y sistematizar criterios objetivos de selección; no obstante, cuando esos modelos se entrena con datos históricos que ya incorporan desigualdades, terminan reproduciéndolas y escalándolas. Lejos de neutralizar los sesgos humanos, la IA puede convertirse en una “máquina de amplificación” de patrones históricos de exclusión (Sheard et al., 2025).

En este trabajo se adopta el término “discriminación algorítmica facilitada” para designar ese fenómeno: situaciones en las que la infraestructura algorítmica reduce los costos de discriminar, hace más difícil la detección de patrones injustos y confiere a las decisiones un halo de objetividad que las vuelve más defendibles. La discriminación no surge solo de malas intenciones individuales, sino de la combinación de datos sesgados, objetivos de optimización mal diseñados y opacidad técnica.

Un caso paradigmático que ilustra este problema en el ámbito laboral es EEOC v. iTutorGroup. Allí se alegó que la empresa había configurado su sistema automatizado de selección para descartar automáticamente a postulantes por encima de determinada edad, en violación de la normativa antidiscriminatoria. El episodio muestra que los algoritmos pueden operar como canales privilegiados para importar prejuicios al corazón de los procesos de reclutamiento, al tiempo que dificultan su detección y prueba en sede administrativa o judicial.

### 1.2. La Insuficiencia del Paradigma *Homo Economicus* ante la Rationalidad Limitada Codificada

El modelo económico clásico del *homo economicus* se fundamenta en la idea de agentes perfectamente racionales, con información completa y capacidad para maximizar su utilidad de manera consistente. Aplicado a la discriminación, este enfoque sugiere que, en mercados competitivos, los empleadores que discriminan asumirán un costo que los volverá menos eficientes frente a quienes contratan al mejor candidato disponible sin sesgos. A largo plazo, la pre-

sión competitiva debería disciplinar la discriminación, que se vuelve económicamente insostenible.

Sin embargo, los hallazgos de la economía comportamental y la literatura sobre sesgos cognitivos muestran que la conducta real de individuos y organizaciones se aparta de ese ideal. En entornos complejos y bajo presión temporal, las decisiones se toman con información incompleta, mediante reglas de dedo y bajo la influencia de normas sociales y estructuras de poder. Cuando sistemas de IA se entrena con esos datos y reglas, no corrigen la racionalidad limitada: la codifican. Los modelos aprenden patrones de decisiones humanas, incluidos sus sesgos, y los replican a escala, con apariencia de neutralidad matemática.

Desde la perspectiva de esta investigación, esta combinación revela una doble insuficiencia del paradigma clásico. En términos descriptivos, el *homo economicus* no ofrece una representación adecuada de cómo se toman decisiones en mercados laborales mediados por algoritmos, donde la agencia está distribuida entre empleadores, desarrolladores y plataformas. Pero, más importante aún, en términos normativos, la promesa de que la competencia eliminará las preferencias discriminatorias deja de ser realista cuando la discriminación se incorpora a infraestructuras digitales ubicuas y difíciles de auditar. En lugar de disiparse, puede estabilizarse y volverse estructuralmente eficiente para ciertos actores, aun cuando resulte socialmente costosa e injusta a nivel sistémico.

### 1.3 La Brecha Académica Identificada

Existe una laguna crítica en la literatura académica internacional: mientras que la economía laboral ha analizado extensivamente la discriminación laboral tradicional (desde los trabajos seminales de Gary Becker en 1957 hasta los análisis contemporáneos de disparidades raciales), y mientras que el derecho ha comenzado recientemente a abordar gobernanza algorítmica, falta un marco integrado que explique por qué los algoritmos, diseñados teóricamente para ser “objetivos” y “neutrales”, reproducen sistemáticamente las mismas inequidades económicas que la discriminación de preferencia personal genera.

La investigación socio-jurídica de Sheard et al. (2025) identifica correctamente que la discriminación facilitada por algoritmos surge no solo de factores técnicos sino de “un conjunto de actores humanos y no-humanos” donde el algoritmo de aprendizaje automático es solo un elemento de un ensamblaje social más amplio. Sin embargo, esa investigación carece de análisis económico riguroso sobre: (i) los costos sociales específicos que estos sesgos generan; (ii) cómo los algoritmos modifian equilibrios de poder previos en negociaciones laborales; (iii) cuantificación de externalidades negativas no internalizadas por las empresas que despliegan estos sistemas.

### 1.4 Relevancia para Gobernanza Algorítmica

Este análisis contribuye directamente al campo emergente de gobernanza algorítmica al demostrar que la regulación de sistemas de IA en recursos humanos no puede comprenderse sin integrar análisis económico riguroso. La “discriminación algorítmica” representa un caso paradigmático de cómo la IA transforma simultáneamente tanto el derecho (redefiniendo qué constituye una “decisión discriminatoria” en contextos de opacidad computacional) como la economía (generando inequidades de mercado y externalidades negativas específicas).

### 1.5 Objetivos Específicos de Investigación

Este análisis persigue tres objetivos integrados:

- a. Proponer el estudio para el desarrollo de una teoría de responsabilidad algorítmica bajo el derecho de la discriminación laboral que reconozca formas novedosas de exclusión sistemática derivadas no de intención sino de arquitectura de datos.
- b. Objetivo Económico: Proponer el estudio para cuantificar las externalidades negativas de discriminación algorítmica mediante el análisis económico de disparidades salariales, tasas de desempleo y costos transaccionales antes y después de automatización de procesos de selección.
- d. Objetivo Regulatorio: Proponer el estudio para el desarrollo de marcos de gobernanza que combinen auditorías de sesgos (medida técnica) con análisis de impacto económico (medida jurídico-económica), identificando el modelo regulatorio comparativamente más efectivo.

### 1.6 Estructura y Contribución

El trabajo se estructura en tres movimientos analíticos progresivos:

Primer, una revisión crítica del estado del arte que examina cómo la economía laboral clásica, la revolución comportamental y la emergente literatura de IA convergen en puntos de intersección no explorados;

Segundo, un análisis pormenorizado de cómo seis tipologías específicas de sesgos operacionales se manifiesta en sistemas de IA de recursos humanos, con énfasis en mecanismos de amplificación sistemática;

Tercero, el desarrollo de una posible propuesta teórica (nueva) que reconceptualiza la discriminación algorítmica como una institución económica estructural antes que como error técnico aislado.

La principal contribución es conceptual: se demuestra que cualquier intento de abordar sesgos algorítmicos únicamente mediante “auditorías técnicas” de precisión o “limpieza de datos” es insuficiente sin comprensión simultánea de cómo estos sistemas modifican estructuras de poder económico, generan nueva asimetría informativa y crean externalidades que no son internalizadas por quienes despliegan la tecnología.

## 2. REVISIÓN DE LITERATURA Y ESTADO DEL ARTE

### 2.1 Discriminación Económica: Teoría Clásica (Becker, Arrow, Phelps)

La teoría económica clásica de la discriminación, iniciada con el trabajo (que se basó en su tesis doctoral) seminario de Gary Becker en 1957 (*The Economics of Discrimination*), plantea que la discriminación en mercados laborales es económicamente irracional: un empleador que rechaza trabajadores productivos por características no económicamente relevantes (raza, género, edad) incurre en costos de oportunidad significativos y está dejando de usar talento disponible y productivo porque sus “preferencias discriminatorias” lo llevan a una elección subóptima.

La predicción teórica clásica de Becker es que, en mercados verdaderamente competitivos, esta discriminación debería ser eliminada gradualmente por presiones competitivas: empresas que no discriminan tendrían acceso a un pool más amplio de talento potencial, reducirían costos laborales, y eventualmente desplazarían a empresas discriminadoras del mercado. Sin embargo, Becker reconoció la existencia de “discriminación de gusto” (*taste-based discrimination*) que podría persistir si los costos de discriminación fueran suficientemente bajos.

Kenneth Arrow (1971) expandió este análisis introduciendo la noción de “discriminación estadística” de empleadores que no tienen información completa sobre productividad individual pueden usar características demográficas observables (grupo racial, género, edad) como proxy estadístico de productividad esperada. Incluso si esta correlación es imprecisa, la discriminación estadística puede ser “racional” desde la perspectiva del empleador que opera bajo incertidumbre.

Edmund Phelps (1972) formalizó que, en contextos de información incompleta, los estereotipos estadísticos pueden persistir como equilibrio incluso si las creencias subyacentes son factualmente incorrectas. Un fenómeno que denomina “racismo estadístico” pero que aplica a cualquier forma de discriminación basada en características de grupo.

Laguna teórica identificada: La literatura de Becker-Arrow-Phelps asume que preferencias discriminatorias son exógenas (provienen de fuera del sistema económico, son preexistentes). Además, asume que discriminación requiere una decisión discreta y consciente por parte del empleador. Los algoritmos de aprendizaje automático generan discriminación de manera endógena—el sistema “aprende” a discriminar del análisis estadístico de datos históricos que contienen discriminación previa. No hay preferencia consciente del empleador, no hay decisión discreta. Hay un proceso de optimización matemática que reproduce patrones previos.

### 2.2 La Revolución Comportamental: Racionalidad Limitada y Sesgos Cognitivos

La revolución comportamental en economía, encabezada por Daniel Kahneman y Amos Tversky (1974), demostró empíricamente que los seres humanos no toman decisiones conforme al modelo de “homo economicus” racional. En su lugar, utilizamos “heurísticos” atajos mentales que simplifican problemas complejos pero que generan “sesgos” sistemáticos en nuestras decisiones.

Kahneman y Tversky identificaron docenas de sesgos cognitivos específicos: sesgo de disponibilidad (recordamos eventos recientes más fácilmente), sesgo de anclaje (nos vemos influidos excesivamente por la primera información que recibimos), sesgo de confirmación (buscamos información que confirma nuestras creencias previas), entre otros.

En el contexto de decisiones de selección laboral, Richard Thaler (2015) documentó cómo incluso evaluadores bien intencionados cometen errores sistemáticos: permitimos que una característica sobresaliente (efecto halo) influencie toda nuestra evaluación de una persona; damos peso excesivo a eventos recientes (sesgo de memoria); valoramos más a quienes percibimos como similares a nosotros (sesgo de similitud).

Implicación crítica para algoritmos: Los algoritmos fueron diseñados precisamente para eliminar estos sesgos cognitivos humanos. Si la máquina toma decisiones basada en cálculo matemático puro, sin influencias emocionales, sin atajos mentales—¿no debería haber menos discriminación, no más?

La respuesta que explora este análisis es: *los algoritmos no eliminan sesgos, los transforman*. Un sesgo cognitivo humano es un patrón en decisiones individuales. Un sesgo algoritmo es un patrón sistemático, ampliado a escala masiva, codificado en matemática, y extraordinariamente difícil de detectar.

### 2.3. Asimetría de la Información y Fallos de Mercado: Del “Mercado de Limones” de Akerlof a la “Caja Negra” Algorítmica

La teoría de la “información asimétrica”, formalizada por George Akerlof (1970) en su influyente trabajo sobre el “mercado de los limones” (*The Market for ‘Lemons’*), es fundamental para comprender por qué los mercados fallan (5). Akerlof demostró que cuando los vendedores tienen más información sobre la calidad de un producto (por ejemplo, un coche usado, que puede ser un “melocotón” o un “limón”) que los compradores, se produce un fenómeno de selección adversa. Los compradores, incapaces de distinguir la calidad, solo están dispuestos a pagar un precio promedio, lo que provoca que los vendedores de productos de alta calidad se retiren del mercado. En casos extremos, el mercado puede colapsar por completo, con solo los “limones” de baja calidad siendo intercambiados (5).

El mercado laboral es un ejemplo clásico de asimetría de información: los candidatos poseen información privada sobre sus verdaderas habilidades y potencial, mientras que los empleadores deben basar sus decisiones en señales imperfectas como la educación o la

experiencia previa. Las herramientas de IA se presentan precisamente como una solución a este problema, prometiendo ser “máquinas de predicción” superiores, capaces de analizar vastas cantidades de datos para identificar a los mejores candidatos y reducir la incertidumbre del empleador (6).

Sin embargo, estos sistemas introducen una asimetría de información de segundo orden: la opacidad de la “caja negra” algorítmica (1). Ni el candidato rechazado ni, con frecuencia, el propio gerente de recursos humanos que utiliza la herramienta comprende los criterios, ponderaciones y correlaciones que el algoritmo utilizó para llegar a una decisión (1). Esta falta de transparencia y explicabilidad impide la verificación, la corrección de errores y la rendición de cuentas, convirtiendo al propio algoritmo en un “limón”: una herramienta de calidad desconocida cuyas decisiones no pueden ser plenamente confiadas (7). Esta opacidad cualitativamente distinta a otras formas de secreto institucional agrava los fallos del mercado en lugar de resolverlos (1).

## 2.4. Perspectivas Críticas y Sociotécnicas sobre el Poder Algorítmico

Un análisis completo de la discriminación algorítmica requiere integrar las aportaciones de la teoría crítica de la tecnología y de los estudios sociotécnicos. Autoras como Safiya Noble han mostrado que los sistemas de información no son meros espejos neutrales de la realidad, sino artefactos que incorporan y refuerzan estructuras de poder preexistentes. Desde esta perspectiva, los algoritmos que ordenan, clasifican y priorizan información operan sobre datos ya atravesados por relaciones de raza, género y clase, y tienden a naturalizar esas jerarquías bajo la apariencia de resultados “objetivos” o “meramente técnicos”.

De manera complementaria, estudios empíricos como Gender Shades evidencian que el desempeño de los sistemas de reconocimiento facial y clasificación de género no es homogéneo entre grupos. En varios servicios comerciales analizados, las tasas de error resultan significativamente más altas para mujeres de piel más oscura que para hombres de piel clara, lo que ilustra cómo la combinación de conjuntos de datos sesgados y objetivos de diseño poco reflexivos puede producir daños desproporcionados sobre grupos ya vulnerabilizados. Este tipo de hallazgos subraya que el problema no se limita a errores aleatorios: los sesgos algorítmicos se apoyan en estructuras de desigualdad históricas y tienden a reproducirlas salvo que se intervenga deliberadamente sobre datos, métricas y formas de evaluación.

## 2.5 Sesgos Algorítmicos: Literatura Emergente 2020-2025

La literatura reciente sobre sesgos algorítmicos ha convergido en ciertos puntos de consenso (Sheard et al., 2025; Coglianese & Lehr, 2018):

1. Sesgos que no surgen solo de “mala programación”, sino de cómo se define qué es un “buen” resultado (función de optimización), qué datos se usan para entrenar (datos históricos que incorporan discriminación previa), y cómo se miden resultados (métricas que pueden no capturar equidad).
2. “Limpieza de datos” es insuficiente: Algunos sugieren que si simplemente se “remueven” variables sensibles (raza, género) del dataset de entrenamiento, se eliminan sesgos. La evidencia muestra que “discriminación por proxy” ocurre cuando variables aparentemente neutrales (código postal, patrón de educación, historial laboral) están altamente correlacionadas con características protegidas.
3. Búcles de retroalimentación amplifican sesgos: Si un algoritmo descalifica desproporcionadamente a ciertos grupos demográficos, estos grupos quedan subrepresentados en futuras contrataciones, futuros datos históricos contienen menos ejemplos de estos grupos en “desempeño exitoso”, el algoritmo aprende nuevas discriminaciones, y el ciclo se auto-refuerza.
4. Opacidad computacional como característica, no fallo: Los sistemas de aprendizaje profundo (deep learning) son interpretables solo de manera limitada incluso por sus diseñadores. Esta opacidad no es un defecto accidental sino una característica arquitectónica. Tiene implicaciones legales profundas: ¿cómo responsabilizar a alguien por discriminación si ni siquiera quienes crearon el sistema pueden explicar exactamente qué está haciendo?

## 3. FUNDAMENTOS TEÓRICOS JURÍDICO-ECONÓMICOS

### 3.1 Teoría de Discriminación Económica (Becker 1957; Arrow 1971; Phelps 1972)

Gary Becker (1957) demostró que empleadores con “preferencias discriminatorias” pueden sostener discriminación si: (i) el costo de discriminación es bajo; (ii) existen barreras a entrada de competidores no discriminadores; (iii) los clientes también tienen preferencias discriminatorias.

Aplicación a algoritmos: Los algoritmos reducen dramáticamente el costo de discriminación. En contratación tradicional, discriminar requería decisiones humanas, que pueden ser escrutinizadas, que pueden ser cuestionadas en juicio, que dejan rastros de evidencia. Un algoritmo discrimina a escala masiva (millones de candidatos), automáticamente (sin interventor humano que pueda testificar sobre intención), con apariencia de objetividad (matemática pura), dejando pocas pruebas de la discriminación ocurrida.

### 3.2 Información Asimétrica en Mercados Laborales (Akerlof 1970; Spence 1973)

George Akerlof (1970) en “The Market for Lemons” (ya se amplió en el punto 2) demostró que cuando existe información asimétrica significativa (una parte tiene información que la otra no tiene), los mercados pueden experimentar ineficiencias severas, incluyendo mercados que potencialmente “desaparecen” completamente.

Michael Spence (1973) desarrolló la teoría de “job market signaling” en presencia de información incompleta sobre productividad de trabajadores, ambas partes (empleador y empleado) gastan re-

cursos en transmitir o verificar información (credenciales, referencias, entrevistas).

### **3.3 Aplicación a algoritmos: Los algoritmos introducen una capa Adicional de asimetría informatacional**

- Ni los candidatos ni frecuentemente los propios empleadores comprenden exactamente cómo el sistema algorítmico toma decisiones.
- La “caja negra” computacional es opaca incluso a análisis especializado.
- Esto viola el principio de “informed consent” los candidatos no saben realmente a qué están siendo evaluados.

En consecuencia, el sesgo algorítmico puede definirse como la presencia de discriminación o prejuicio en la representación de la realidad que influye en la toma de decisiones o en los resultados de un sistema de IA.

Numerosos estudios han identificado percepciones subconscientes comunes que influyen en la evaluación del rendimiento. No todos los prejuicios se originan en la discriminación explícita. A continuación, enumeraremos algunos de los tipos de sesgos que podemos ver presentes en los procesos de contratación y evaluación del desempeño.

En primer lugar, nos encontramos con el sesgo de memoria, que implica que cuanta más distancia hay entre un hecho y su recuerdo, mayor es la dificultad para recordarlo con precisión. En una evaluación de desempeño, este sesgo implica dar más peso a los acontecimientos recientes y descartar o minimizar lo sucedido antes.

Por su parte, el sesgo de proximidad hace referencia al hecho de valorar más a quienes se encuentran con frecuencia cerca. Con el auge del trabajo remoto e híbrido, este sesgo podría afectar especialmente a quienes trabajan desde casa, independientemente de su rendimiento o compromiso real.

El sesgo de primacía consiste en permitir que las primeras impresiones influencien las evaluaciones posteriores de personalidad, habilidades y otras competencias, pasando por alto información relevante que surja más adelante. Este sesgo opera en dos direcciones: una buena primera impresión puede hacer que se desestimen comportamientos posteriores menos positivos; mientras que una mala primera impresión puede teñir de manera más crítica las acciones futuras.

Luego, tenemos el efecto halo que, básicamente, es la tendencia a que un atributo destacado, ya sea positivo o negativo, tiña toda la evaluación de una persona. Por ejemplo, podemos pensar en el supuesto de que un único logro sobresaliente puede ocultar deficiencias en colaboración o trabajo en equipo.

Por último, podemos mencionar al sesgo de similitud que parte de la premisa de que los seres humanos tienden a sentir afinidad por quienes perciben como similares. En virtud de ello, en evalua-

ciones de rendimiento, esto se manifiesta o traduce en una preferencia inconsciente por valorar más a quienes permiten verse reflejado en ellos.

En efecto, entre los riesgos más significativos para las empresas al implementar la IA, hallamos la posibilidad de obtener resultados sesgados. Si este sesgo se proyecta sobre datos personales, puede generar discriminación o sesgos injustos basados en sexo, raza, origen, creencia u otras características, con impactos negativos en grupos de personas y en la reputación de la empresa.

El sesgo en IA se manifiesta como un fallo sistemático que surge cuando se favorece ciertas respuestas sobre otras. Para entender sus causas, es clave revisar conceptos fundamentales de los modelos de IA.

### **3.4 Aplicación a algoritmos de IA**

La implementación de sistemas de IA en selección laboral suele presentarse como un intento de reducir costos transaccionales:

- Costo de buscar candidatos cualificados: reducido (más candidatos en pool)
- Costo de evaluar candidatos: reducido (máquina evalúa automáticamente)
- Costo de negociación: potencialmente aumentado (candidatos menos información sobre cómo son evaluados)
- Costo de resolución de disputas: aumentado (opacidad algorítmica dificulta reclamos de discriminación)

La pregunta económica es: ¿El algoritmo logra reducción neta de costos transaccionales, o simplemente traslada costos hacia trabajadores y hacia sistema legal?

### **3.5 Jurisprudencia Clave: Análisis Profundo de EEOC v. iTutor-Group y Mobley v. Workday, Inc.**

La jurisprudencia está comenzando a aplicar estos principios al contexto algorítmico, con dos casos que destacan por sus implicaciones.

El caso *EEOC v. iTutorGroup* representa un ejemplo claro de trato dispar mediado por un algoritmo. La empresa programó explícitamente su software para descartar a solicitantes basándose en su fecha de nacimiento, estableciendo umbrales de edad diferentes para hombres y mujeres (3). El acuerdo resultante, que incluyó una compensación monetaria de \$365,000 y medidas cautelares como la prohibición de solicitar la fecha de nacimiento de los solicitantes, estableció un precedente crucial: las empresas son directamente responsables del código discriminatorio que implementan, incluso si el proceso es totalmente automatizado (3).

El caso *Mobley v. Workday, Inc.* es considerablemente más complejo y de mayor alcance, ya que aborda la teoría del impacto dispar en el contexto de un proveedor de tecnología de terceros (1). El de-

mandante, Derek Mobley, alega que las herramientas de selección algorítmica de Workday, utilizadas por miles de empresas, discriminan sistemáticamente a los solicitantes afroamericanos, mayores de 40 años y con discapacidades (12). Una decisión judicial clave en este caso fue la certificación preliminar de una acción colectiva a nivel nacional bajo la ADEA, permitiendo que otros solicitantes mayores de 40 años se unan a la demanda (12). El tribunal determinó que la cuestión central es si el sistema de IA de Workday tiene un impacto dispar en los solicitantes mayores de 40 años es una cuestión común que puede abordarse colectivamente, sentando un precedente importante para futuras demandas colectivas contra sistemas de IA (11).

#### 4. HACIA UN MODELO INTEGRADO: CONCLUSIONES Y RECOMENDACIONES POLÍTICO-NORMATIVAS

##### 4.1. Síntesis de Hallazgos: La Discriminación Algorítmica como un Fallo de Mercado Institucionalizado

Esta investigación ha demostrado que los sesgos algorítmicos en la contratación no son meros errores técnicos ni simples violaciones legales aisladas. Constituyen una manifestación de un fallo de mercado sistémico, institucionalizado a través de la tecnología. Al integrar la teoría económica de la discriminación, la economía de la información y la economía de los costos de transacción con el análisis jurídico de la jurisprudencia y la regulación emergente se llega a una conclusión central: los algoritmos de contratación, al aprender de datos históricos sesgados, crean una forma endógena de discriminación que es económicamente ineficiente. Reducen los costos de transacción de discriminar a escala, amplifican las asimetrías de información a través de su opacidad y generan externalidades negativas en forma de mala asignación de talento y pérdida de bienestar social. Los marcos regulatorios actuales, aunque bien intencionados, abordan los síntomas (la falta de equidad estadística) sin atacar la causa económica subyacente (la ineficiencia del mercado).

##### 4.2. Propuesta de un Marco Regulatorio Dual: Integración Obligatoria de Auditorías de Sesgo y Análisis de Impacto Económico

Para superar las limitaciones de los enfoques regulatorios existentes, se propone un marco regulatorio dual que integre de manera inseparable el análisis jurídico y el económico (1). Este modelo requeriría que las empresas no solo cumplan con las auditorías de sesgos, sino que también realicen y publiquen un análisis de impacto económico.

**1. Auditoría de Sesgos (Componente Jurídico):** Siguiendo el modelo de la Ley Local 144 de Nueva York, se debe exigir una auditoría de impacto dispar realizada por un tercero independiente antes del despliegue de cualquier sistema de IA de alto riesgo en la contratación. Sin embargo, estas auditorías deben estar

estandarizadas, requiriendo análisis interseccionales (como los demostrados por el estudio “Gender Shades”) y no solo evaluaciones de variables únicas (10).

**2. Análisis de Impacto Económico (Componente Económico):** La novedad de esta propuesta radica en exigir que la auditoría se complementa con una estimación del impacto económico de la herramienta. Esto implicaría que la empresa debe cuantificar la tasa de “falsos negativos” (candidatos cualificados rechazados) dentro de los grupos demográficos subrepresentados. La publicación de este análisis obligaría a las empresas a internalizar y transparentar la ineficiencia económica que sus propias herramientas generan, desplazando el debate desde la mera equidad estadística hacia la optimización del talento y la eficiencia del mercado.

#### 4.3. Recomendaciones para la Expansión de la Responsabilidad Jurídica bajo un Régimen de Responsabilidad Objetiva

La evolución de la doctrina jurídica, especialmente la aceptación de la teoría del “agente” en el caso *Mobley v. Workday*, abre la puerta a un régimen de responsabilidad más eficaz (13). Se recomienda la codificación legislativa de una forma de responsabilidad objetiva (*strict liability*) para los desarrolladores de sistemas de IA de alto riesgo utilizados en el empleo (1).

El argumento económico para esta recomendación se basa en el principio del “evitador de menor costo” (*least-cost avoider*). El desarrollador del algoritmo, no el empleador que lo utiliza está en la mejor posición para prevenir, detectar y mitigar los sesgos durante las fases de diseño, entrenamiento y validación del modelo. Un régimen de responsabilidad objetiva crearía el incentivo económico más fuerte posible para que los proveedores inviertan en el desarrollo de tecnologías justas y precisas desde el principio, ya que no podrían simplemente transferir el riesgo legal a sus clientes. Esto alinearía sus intereses comerciales con el objetivo social de un mercado laboral eficiente y equitativo.

#### 5. METODOLOGÍA

Este trabajo se estructura como un estudio teórico jurídico-económico que propone, además, un diseño empírico para evaluar la existencia de discriminación algorítmica e ineficiencias en mercados de trabajo. No se reportan resultados cuantitativos: se presenta un plan metodológico replicable que puede ser implementado por organizaciones o equipos de investigación con acceso a los datos pertinentes.

El enfoque combina dos dimensiones. Por un lado, un análisis conceptual y normativo de la discriminación algorítmica, basado en la teoría económica de la discriminación, la literatura sobre información asimétrica y los desarrollos recientes en torno a sesgos algorítmicos. Por otro lado, se propone un módulo econométrico orientado a medir impacto dispar en procesos de selección y brechas salariales entre grupos protegidos, utilizando datos observacionales.

En términos de datos, el diseño prevé dos escenarios. En el primero (A), se trabaja con registros internos de recursos humanos de una organización que utiliza herramientas algorítmicas en reclutamiento y gestión de personal, lo que permite vincular decisiones automatizadas con características de los postulantes y resultados laborales. En el segundo (B), cuando esos registros no están disponibles, se recurre a encuestas u otras fuentes externas que permitan aproximar patrones de selección, remuneración y permanencia. Si la organización no facilita datos internos, el escenario A se mantiene como plan, y se implementa el escenario B con información pública o de encuestas representativas.

Las variables dependientes propuestas son: (i) un indicador de selección o contratación (seleccionado/no seleccionado) y (ii) el logaritmo del salario horario o mensual. Como variables explicativas principales se utilizan indicadores de pertenencia a grupos protegidos según la normativa aplicable (por ejemplo, género, edad, origen étnico, discapacidad), y como controles se incorporan nivel educativo, experiencia, sector, tamaño de firma, canal de postulación, ocupación y región, en la medida en que la información esté disponible.

Para evaluar la equidad en selección se calcula la tasa de contratación por grupo y la denominada “regla del cuatro quintos” (ratio entre la tasa de selección del grupo protegido y la del grupo de referencia). En el plano salarial, se estiman brechas ajustadas comparando el salario esperado de grupos protegidos y no protegidos, una vez controlados los factores relevantes.

El modelo estadístico propuesto se organiza en dos etapas. En la primera, se estima una regresión logística de la probabilidad de ser seleccionado, con efectos fijos por ocupación y región, reportando razones de odds y efectos marginales para cada grupo protegido. En la segunda, se estima un modelo de regresión lineal sobre el logaritmo del salario, con los mismos efectos fijos, reportando coeficientes, intervalos de confianza y  $R^2$ , e interpretando los coeficientes en términos porcentuales.

La validez de las estimaciones se discute mediante análisis de sensibilidad, agrupando errores por vacante u organización, y evaluando la robustez de los resultados al excluir posibles variables que actúen como proxies de características protegidas. Se enfatiza que, al tratarse de datos observacionales, no se infiere causalidad fuerte sin diseños quasi-experimentales adicionales.

Finalmente, el plan metodológico incorpora criterios de ética y cumplimiento. El uso de atributos sensibles se justifica exclusivamente para la evaluación de sesgos y con resguardo estricto de la

confidencialidad. Se exige documentar las transformaciones de datos, el software utilizado y cualquier restricción de acceso, de modo que otros equipos puedan reproducir los análisis. Cuando no se cuente con información interna suficiente, se explicitan las limitaciones y se desaconseja extraer conclusiones generalizadas sin evidencia complementaria.

## 6. POSIBLES FUTURAS LÍNEAS DE INVESTIGACIÓN: EL ROL DE LA TRANSPARENCIA EXPLICABLE (XAI) Y LA GOBERNANZA DE DATOS

Aunque el modelo regulatorio dual propuesto —auditorías de sesgos e integración sistemática de análisis de impacto económico— ofrece una base sólida para abordar la discriminación algorítmica, deja abiertas varias líneas de trabajo. En particular, se vuelve central investigar cómo las herramientas de transparencia y explicabilidad pueden integrarse de manera operativa en la supervisión de sistemas de IA aplicados a la gestión del trabajo, y cómo la gobernanza de datos condiciona las posibilidades efectivas de control.

1. Estándares de transparencia y explicabilidad (XAI). Futuras investigaciones deberían analizar qué técnicas de explicabilidad son más adecuadas para contextos laborales específicos (por ejemplo, procesos de screening masivo frente a promociones internas) y cómo traducir sus resultados en insumos comprensibles para personas trabajadoras, sindicatos, autoridades administrativas y jueces. No se trata solo de producir explicaciones técnicas, sino de construir mecanismos institucionales que permitan cuestionar decisiones automatizadas cuando producen impactos discriminatorios.

2. Gobernanza de datos y documentación estandarizada. Otra línea clave es el estudio de marcos de gobernanza de datos que garanticen calidad, representatividad y trazabilidad en los conjuntos utilizados para entrenar y validar modelos aplicados a recursos humanos. Instrumentos como las “model cards” y las fichas de documentación de conjuntos de datos pueden funcionar como piezas centrales de una gobernanza algorítmica que promueva transparencia, responsabilidad y participación de los grupos afectados.

La integración de estas herramientas técnicas en un marco de gobernanza más amplio, basado en derechos fundamentales y en principios de no discriminación, puede contribuir a que la adopción de IA en los mercados de trabajo no se logre a costa de la equidad y la justicia económica, sino que se oriente a corregir desigualdades históricas y a mejorar de manera verificable el bienestar de quienes trabajan.

## 7. REFERENCIAS BIBLIOGRÁFICAS

- Acemoglu, D., & Restrepo, P. (2023). The Fall of the Labor Share and the Rise of Superstar Firms. *Journal of Political Economy*, 130(2), 1-30. <https://doi.org/10.1086/721816>
- Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Review Press.
- Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3), 488-500. <https://doi.org/10.2307/1879431>
- Arrow, K. J. (1971). The Theory of Discrimination. Working Paper. National Bureau of Economic Research. <https://www.nber.org/papers/w0018>
- Becker, G. S. (1957). The Economics of Discrimination. University of Chicago Press.
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the Conference on Fairness, Accountability and Transparency*, 77–91.
- Kahneman, D., & Tversky, A. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>
- Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. New York University Press.
- NYC Local Law 144. (2023). Automated Employment Decision Tools Law. Effective January 1, 2023. City of New York.
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, 62(4), 659–661.
- Sheard, L., Lautenbach, F., Lace, S., & Wischmann, S. (2025). Algorithm-Facilitated Discrimination. *Journal of Law and Society*. <https://doi.org/10.1111/jols.12535>
- Spence, M. (1973). Job market signaling. *The Bell Journal of Economics*, 4(2), 355–374. <https://doi.org/10.2307/3003186>
- Thaler, R. H. (2015). Misbehaving: The Making of Behavioral Economics. W.W. Norton & Company.
- Wilson, V., & Darity Jr., W. (2021). Understanding Black-White Disparities in Labor Market Outcomes: Requires Models That Account for Persistent Discrimination and Unequal Bargaining Power. Economic Policy Institute. <https://www.epi.org/unequalpower/publications/understanding-black-white-disparities-in-labor-market-outcomes/>
1. "Sesgos y discriminaciones sociales de los algoritmos en Inteligencia Artificial: una revisión documental" acceso: octubre 19, 2025, <https://dialnet.unirioja.es/servlet/articulo?codigo=9380872>
2. US EEOC's first settlement in AI hiring discrimination | United States - Norton Rose Fulbright, acceso: octubre 19, 2025, <https://www.nortonrosefulbright.com/en-us/knowledge/publications/2ec12415/us-eecs-first-settlement-in-ai-hiring-discrimination>
3. Equal Employment Opportunity Commission v. iTutorGroup, Inc. 1 ..., acceso: octubre 05, 2025, <https://clearinghouse.net/case/44258/>
4. Judgment under Uncertainty: Heuristics and Biases - PubMed, acceso: octubre 05, 2025, <https://pubmed.ncbi.nlm.nih.gov/17835457/>
5. The Market for "Lemons": Quality Uncertainty and the Market ..., acceso: octubre 05, 2025, <https://www.sfu.ca/~wainwrig/Econ400/akerlof.pdf>
6. Prediction Machines, Updated and Expanded: The Simple Economics of Artificial Intelligence - HBR Store, acceso: octubre 05, 2025, <https://store.hbr.org/product/prediction-machines-updated-and-expanded-the-simple-economics-of-artificial-intelligence/10598>
7. "Why Should I Trust You?": Explaining the Predictions of Any Classifier - ResearchGate, acceso: octubre 05, 2025, [https://www.researchgate.net/publication/305999024\\_Why\\_Should\\_I\\_Trust\\_You\\_Explaining\\_the\\_Predictions\\_of\\_Any\\_C](https://www.researchgate.net/publication/305999024_Why_Should_I_Trust_You_Explaining_the_Predictions_of_Any_C)
8. Algorithms of Oppression: How Search Engines Reinforce Racism - Project MUSE, acceso: octubre 05, 2025, <https://muse.jhu.edu/book/64995>
9. Gender Shades, acceso: octubre 05, 2025, <http://gendershades.org/overview.html>
10. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, acceso: octubre 05, 2025, <https://proceedings.mlr.press/v81/buolamwini18a.html>
11. UNITED STATES DISTRICT COURT NORTHERN DISTRICT OF CALIFORNIA DEREK MOBLEY, Plaintiff, v. WORKDAY, INC., Defendant. Case No. 2 - GovInfo, acceso: octubre 05, 2025, [https://www.govinfo.gov/content/pkg/USCOURTS-cand-3\\_23-cv-00770/pdf/USCOURTS-cand-3\\_23-cv-00770-1.pdf](https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_23-cv-00770/pdf/USCOURTS-cand-3_23-cv-00770-1.pdf)
12. Federal Court Allows Collective Action Lawsuit Over Alleged AI ..., acceso: octubre 05, 2025,
13. Mobley v. Workday: Court Holds AI Service Providers Could Be ..., acceso: octubre 05, 2025, <https://www.seyfarth.com/news-insights/mobley-v-workday-court-holds-ai-service-providers-could-be-directly-liable-for-employment-discrimination-under-agent-theory.html>

14. Model Cards for Model Reporting - arXiv, acceso: octubre 05, 2025, <https://arxiv.org/pdf/1810.03993>

# AmeliCA

## Disponible en:

<https://portal.amelica.org/ameli/ameli/journal/966/9665479011/9665479011.pdf>

Cómo citar el artículo

Número completo

Más información del artículo

Página de la revista en [portal.amelica.org](https://portal.amelica.org)

AmeliCA

Ciencia Abierta para el Bien Común

Carolina Pinasco

**Algoritmos Discriminadores y Mercados de Trabajo  
Ineficientes: Sesgos y discriminación en la contratación y  
evaluación de desempeño laboral. La Contratación bajo  
el paradigma Post-Homo economicus**

**Discriminatory Algorithms and Inefficient Labor Markets:  
Biases and Discrimination in Hiring and Performance  
Evaluation. Hiring under the Post-Homo Economicus  
Paradigm**

**Algoritmos Discriminatórios e Mercados de Trabalho  
Ineficientes: Vieses e Discriminação na Contratação e  
Avaliação de Desempenho. Contratação sob o Paradigma  
Pós-Homo Economicus**

*Revista Fundamentos*  
núm. 2, 2025

Universidad Nacional de Río Cuarto, Argentina  
[fundamentos@fce.unrc.edu.ar](mailto:fundamentos@fce.unrc.edu.ar)

**ISSN-E:** 2545-6318

**DOI:** <https://doi.org/10.63207/a84d1744>