

Ética y Responsabilidad Social en la Implementación de Sistemas de Inteligencia Artificial en la Administración

Ethics and Social Responsibility in the Implementation of Artificial Intelligence Systems in the Administration

Samuel Isaías Acevedo Torres

<https://orcid.org/0000-0002-1505-6117>

RESUMEN

Esta revisión sistemática tuvo como objetivo analizar los dilemas éticos y las implicaciones de la responsabilidad social en la adopción de la inteligencia artificial (IA) en la administración empresarial. Se realizó una búsqueda exhaustiva en las bases de datos Scopus, Web of Science y EBSCOhost, identificando estudios publicados entre 2018 y 2024 que cumplieran con los criterios de inclusión. La síntesis narrativa de los hallazgos reveló preocupaciones clave en torno a la equidad en la toma de decisiones automatizada, la necesidad de transparencia y explicabilidad de los algoritmos, la definición de la responsabilidad por las acciones de la IA, el impacto en el empleo y las habilidades de los trabajadores, y la protección de la privacidad y los datos. La revisión concluye que la integración ética de la IA en la administración requiere la adopción de directrices éticas claras, marcos de gobernanza robustos y un enfoque proactivo por parte de las empresas para mitigar los riesgos éticos y promover un entorno de trabajo justo, transparente y sostenible.

Palabras clave:

Ética, responsabilidad social, inteligencia artificial, administración empresarial, gestión, toma de decisiones, transparencia, sesgo, privacidad, empleo.

ABSTRACT

This systematic review aimed to analyze the ethical dilemmas and social responsibility implications of artificial intelligence (AI) adoption in business administration. A comprehensive search was conducted in Scopus, Web of Science, and EBSCOhost databases, identifying 15 studies published between 2018 and 2024 that met the inclusion criteria. The narrative synthesis of findings revealed key concerns regarding fairness in automated decision-making, the need for algorithmic transparency and explainability, defining accountability for AI actions, the impact on employment and worker skills, and the protection of privacy and data. The review concludes that ethical AI integration in administration requires the adoption of clear ethical guidelines, robust governance frameworks, and a proactive approach by companies to mitigate ethical risks and promote a fair, transparent, and sustainable work environment.

Keywords:

Ethics, social responsibility, artificial intelligence, business administration, management, decision-making, transparency, bias, privacy, employment.

INTRODUCCIÓN

La inteligencia artificial (IA) está transformando rápidamente el panorama empresarial, impulsando una nueva ola de innovación y eficiencia en la administración. Desde la automatización de tareas rutinarias hasta la toma de decisiones estratégicas, la IA ofrece un potencial sin precedentes para optimizar las operaciones, mejorar la productividad y crear nuevas oportunidades de negocio. Sin embargo, a medida que las organizaciones adoptan la IA a un ritmo acelerado, surgen cuestionamientos cruciales sobre las implicaciones éticas y sociales de delegar decisiones y responsabilidades a sistemas autónomos.

La gestión empresarial, tradicionalmente guiada por principios éticos y marcos legales, se enfrenta ahora al desafío de integrar la IA de manera responsable, asegurando que su implementación se alinee con los valores humanos, la justicia social y el bienestar colectivo. La capacidad de la IA para procesar grandes volúmenes de datos,

identificar patrones complejos y tomar decisiones a una velocidad sin precedentes plantea dilemas éticos complejos.

Uno de los desafíos más apremiantes es la equidad en la toma de decisiones automatizada. Los algoritmos de IA, aunque aparentemente objetivos, pueden perpetuar y amplificar los sesgos existentes en los datos de entrenamiento, generando resultados discriminatorios y perpetuando desigualdades sociales (Rosemann & Zhang, 2022). La falta de transparencia en los procesos de toma de decisiones algorítmicas, a menudo descrita como el problema de la “caja negra”, dificulta la identificación de sesgos, la explicación de resultados y la rendición de cuentas por las consecuencias negativas (Harlow, 2018).

Además de la equidad, la transparencia, la rendición de cuentas y la privacidad emergen como temas centrales en el debate sobre la ética de la IA en la administración. La recolección y el análisis masivo de datos, si bien esenciales para el funcionamiento de la IA, plantea preocupaciones sobre la privacidad individual, el uso indebido de información personal y la vigilancia constante (Santoni de Sio & Mecacci, 2021).

En este contexto, las organizaciones se enfrentan al desafío de gestionar la responsabilidad por las decisiones y acciones tomadas por los sistemas de IA. Definir la responsabilidad en sistemas autónomos, donde la toma de decisiones no siempre puede atribuirse directamente a un individuo o grupo específico, es fundamental para construir confianza, garantizar la justicia y mitigar los riesgos potenciales (Djeffal, 2020).

Abordar estos desafíos requiere un enfoque proactivo y estratégico que integre la ética en el diseño, desarrollo, implementación y gestión de sistemas de IA en el ámbito empresarial. La creación de una cultura ética sólida dentro de las organizaciones, la formación en ética de la IA para los profesionales de la administración y la implementación de marcos de gobernanza robustos son esenciales para garantizar una gestión responsable de la IA.

Este artículo analiza en profundidad los dilemas éticos y las implicaciones de responsabilidad social que surgen con la adopción de la IA en la administración de empresas. A través de una revisión exhaustiva de la literatura académica y casos de estudio, se examinarán los desafíos y oportunidades que la IA presenta para una gestión empresarial ética y socialmente responsable.

El objetivo principal de este artículo es proponer un conjunto de directrices prácticas para la implementación ética y responsable de la IA en las empresas. Estas directrices, basadas en principios de transparencia, responsabilidad, equidad y justicia, buscan guiar a las organizaciones en la integración efectiva de la IA, maximizando sus beneficios y mitigando los riesgos éticos y sociales asociados

MATERIALES Y MÉTODOS

Esta revisión sistemática se ha diseñado siguiendo las directrices PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) para garantizar la transparencia y rigurosidad en la identificación, selección y análisis de estudios sobre las implicaciones éticas y sociales de la inteligencia artificial en la administración empresarial. La adhesión a estas pautas metodológicas ampliamente aceptadas dentro de la comunidad científica refuerza la validez y confiabilidad de los hallazgos de esta revisión (Moher et al., 2009).

Criterios de elegibilidad

Se establecieron criterios de inclusión y exclusión específicos para guiar la selección de estudios relevantes para esta revisión. Tal como muestra la Tabla 2-1:

Tabla 2-1
Criterios de selección

Criterios	Descripción
Inclusión	• Estudios publicados a partir del año 2019 Se priorizaron estudios recientes para reflejar el estado actual del debate sobre ética de la IA, que ha experimentado un desarrollo acelerado en los últimos años.
	• Estudios que aborden la ética de la IA en la administración empresarial Se incluyeron trabajos que analizaran los dilemas éticos, la responsabilidad social, el impacto social y/o la gobernanza de la IA en el contexto específico de las empresas y la gestión empresarial.
	• Estudios con diferentes diseños metodológicos Se consideraron estudios empíricos, revisiones de literatura, análisis de casos, marcos teóricos y directrices éticas para obtener una visión amplia y diversa del tema.
	• Estudios disponibles en inglés o español Se eligieron estos idiomas por ser los de trabajo del investigador.

Criterios	Descripción
Exclusión	Estudios centrados exclusivamente en aspectos técnicos de la IA Se excluyeron estudios que no abordaran las dimensiones éticas o sociales de la IA
	Estudios enfocados en sectores distintos a la administración de empresas Se excluyeron trabajos que se centraran en sectores como salud, educación, derecho, etc., aunque se revisaron para identificar posibles conexiones o aplicaciones al contexto empresarial.
	Estudios no disponibles en texto completo Se excluyeron estudios cuyos textos completos no estuvieran accesibles a través de las bases de datos consultadas o por contacto con los autores.

Fuentes de información

La búsqueda de literatura se realizó en las siguientes bases de datos electrónicas, reconocidas por su amplio alcance en el ámbito de las ciencias sociales y la administración de empresas:

- Scopus:** Reconocida por su cobertura multidisciplinar y su amplio índice de publicaciones académicas, incluyendo revistas de alto impacto.
- Web of Science (WoS):** Base de datos multidisciplinar que incluye las principales revistas científicas del mundo, así como actas de congresos y otros tipos de publicaciones.
- EBSCOhost (Business Source Complete):** Base de datos especializada en negocios y administración, con un enfoque particular en publicaciones académicas y profesionales relevantes para el mundo empresarial.

Estrategia de búsqueda

Se desarrolló una estrategia de búsqueda exhaustiva y precisa para cada base de datos, utilizando una combinación de palabras clave y operadores booleanos para asegurar la recuperación de la mayor cantidad posible de estudios relevantes.

Ecuaciones de búsqueda:

Tabla 2-2

Ecuaciones de búsqueda

Base de Datos	Ecuación de búsqueda
Scopus	TITLE-ABS-KEY (("ethics" OR "ethical") AND ("social responsibility" OR "corporate social responsibility") AND ("artificial intelligence" OR "AI") AND ("administration" OR "management" OR "business")) AND (LIMIT-TO (PUBYEAR, 2019 - 2024))
Web of Science.	ALL = (("ethics" OR "ethical") AND ("social responsibility" OR "corporate social responsibility") AND ("artificial intelligence" OR "AI") AND ("administration" OR "management" OR "business")) AND YEAR PUBLISHED = 2019-2024
EBSCOhost (Business Source Complete)	TX All Text (("ethics" OR "ethical") AND ("social responsibility" OR "corporate social responsibility") AND ("artificial intelligence" OR "AI") AND ("administration" OR "management" OR "business")) AND (YR 2019 TO 2024)

Además de las búsquedas en bases de datos mostradas en la Tabla 2-2, se llevó a cabo una búsqueda manual en las listas de referencias de los artículos seleccionados para identificar estudios potencialmente relevantes que pudieran no haber sido capturados en la búsqueda electrónica inicial (*Snowballing*). Este método, ampliamente utilizado en revisiones sistemáticas, permite ampliar el alcance de la búsqueda y aumentar la exhaustividad de la revisión (Greenhalgh & Peacock, 2005).

Proceso de selección de los estudios

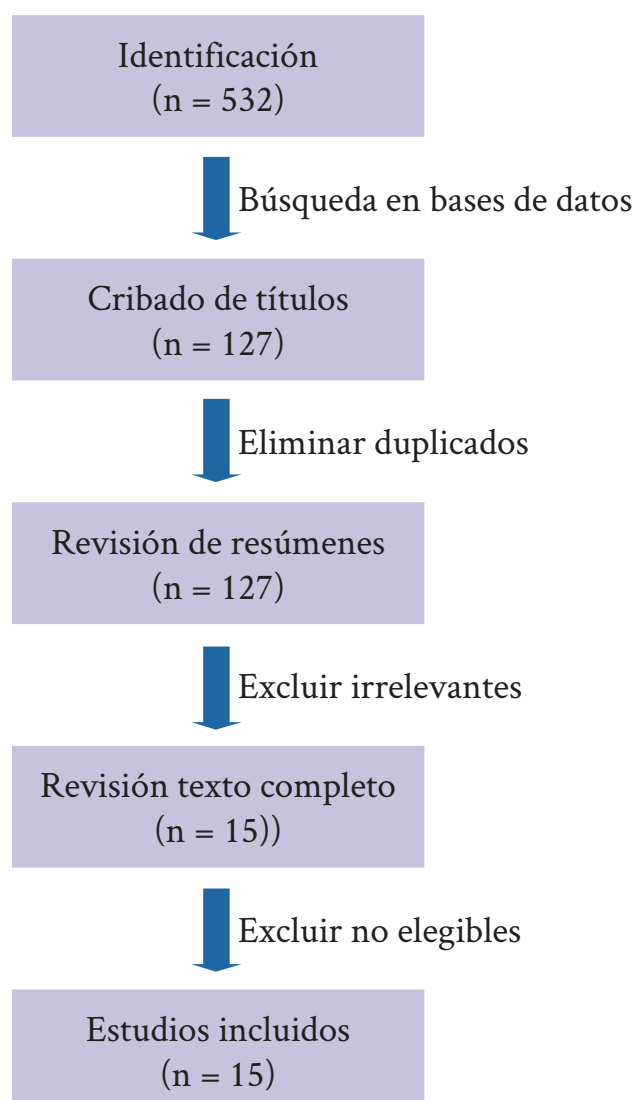
La búsqueda inicial en las bases de datos arrojó un total de 532 registros, incluyendo duplicados. El proceso de selección de los estudios se llevó a cabo en tres etapas, siguiendo una estrategia de doble cribado para minimizar el sesgo y asegurar la inclusión de todos los estudios relevantes:

1. **Revisión de títulos y resúmenes:** En esta etapa inicial, se revisaron los títulos y resúmenes de los 532 registros bibliográficos recuperados en las búsquedas electrónicas y manual. Esta revisión inicial se realizó utilizando un formulario estandarizado para asegurar la consistencia en la aplicación de los criterios de inclusión y exclusión. Los registros que no cumplían claramente con los criterios fueron descartados, resultando en la selección de 127 artículos para la siguiente etapa.

2. **Revisión de texto completo:** Los 127 artículos que potencialmente cumplían con los criterios de inclusión después de la revisión de títulos y resúmenes se recuperaron en texto completo para una evaluación exhaustiva. Se utilizó una plantilla de análisis predefinida para guiar la lectura crítica y la extracción de datos de cada estudio. Tras esta etapa, se descartaron 112 artículos por no cumplir con los criterios de inclusión.
3. **Decisión final:** Después de la revisión en texto completo, se discutieron las discrepancias entre los revisores y se tomó una decisión final sobre la inclusión o exclusión de cada estudio. La revisión completa del texto condujo a la inclusión final de 15 estudios en la revisión sistemática. Los motivos de exclusión de los estudios se documentaron en una tabla para asegurar la transparencia del proceso de selección.

Figura 2-1

Diagrama de flujo del proceso de cribado



Proceso de extracción de los datos

Se diseñó una plantilla de extracción de datos para sistematizar la recopilación de información relevante de cada estudio incluido, considerando las siguientes variables:

- **Información bibliográfica:** Autores, año de publicación, título, revista o editorial, DOI (*Digital Object Identifier*).
- **Diseño del estudio:** Tipo de estudio (empírico, revisión, teórico), metodología utilizada (cualitativa, cuantitativa, mixta).
- **Contexto de la investigación:** Tipo de organización o empresa estudiada (sector, tamaño, ubicación geográfica), características de los participantes.
- **Dilemas éticos abordados:** Descripción de los dilemas éticos identificados en el estudio, enfoques teóricos utilizados para analizarlos.
- **Implicaciones sociales identificadas:** Análisis del impacto social de la IA en la administración (empleo, privacidad, discriminación, etc.).
- **Marcos éticos, directrices o recomendaciones propuestas:** Descripción de las propuestas para la gestión ética de la IA en las empresas (códigos éticos, marcos de gobernanza, principios rectores).
- **Limitaciones del estudio:** Reconocimiento de las limitaciones metodológicas u otras limitaciones relevantes del estudio.

Lista de datos

La información extraída de los 15 estudios seleccionados se organizó en una tabla para facilitar el análisis y la síntesis. La Tabla 2-3 presenta un resumen de los estudios incluidos en la revisión.

Tabla 2-3
Resumen de estudios incluidos sobre ética e IA

Autor(es)	Título	Contexto	Diseño del Estudio
Rosemann & Zhang	<i>Exploring the social, ethical, legal, and responsibility dimensions of artificial intelligence for health-a new column in Intelligent Medicine</i>	Salud, con aplicaciones a la gestión	Editorial/ Reflexión

Autor(es)	Título	Contexto	Diseño del Estudio
Harlow	<i>Ethical concerns of artificial intelligence, big data and data analytics</i>	General, aplicable a la administración	Revisión de literatura
Santoni de Sio & Mecacci	<i>Four responsibility gaps with artificial intelligence: Why they matter and how to address them</i>		Revisión conceptual/ filosófica
Brendel et al.	<i>Ethical management of artificial intelligence</i>		Propuesta de marco teórico
Djeffal	<i>Artificial intelligence and public governance: normative guidelines for artificial intelligence in government and public administration</i>	Administración Pública, aplicable a la gestión empresarial	Análisis de directrices y regulaciones
Weber-Lewerenz	<i>Corporate digital responsibility (CDR) in construction engineering—ethical guidelines for the application of digital transformation and artificial intelligence (AI) in user practice</i>	Ingeniería de construcción, aplicable a la gestión	Revisión de literatura, estudio de casos
Elliott et al.	<i>Towards an equitable digital society: artificial intelligence (AI) and corporate digital responsibility (CDR)</i>	General, aplicable a la administración	Revisión, análisis de casos
Benefo et al.	<i>Ethical, legal, social, and economic (ELSE) implications of artificial intelligence at a global level: a scientometrics approach</i>		Análisis cuantitativo, revisión de literatura
Carter	<i>Regulation and ethics in artificial intelligence and machine learning technologies: Where are we now? Who is responsible? Can the information professional play a role?</i>		Revisión de regulaciones y marcos éticos
Zaman	<i>The weighty responsibility of creating AI navigating control and ethics</i>		Ensayo/ Reflexión
Carrillo	<i>Artificial intelligence: From ethics to law</i>		Análisis jurídico-conceptual
Kerr, Barry & Kelleher	<i>Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance</i>		Análisis sociológico, revisión de documentos

Autor(es)	Título	Contexto	Diseño del Estudio
Kieslich, Keller & Starke	<i>Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence</i>	General, aplicable a la administración	Encuesta, análisis de percepción pública
Nikadimovs & Vēvere	<i>The use of generative artificial intelligence in higher education: university social responsibility and stakeholders' perceptions</i>	Educación Superior, con aplicaciones a la gestión	Estudio piloto, encuesta
Ali et al.	<i>Artificial Intelligence: benefits, application, ethical issues, and organizational responses</i>	General, aplicable a la administración	No especificado

Nota: Elaboración propia a partir de la revisión de la literatura.

Evaluación del riesgo de sesgo de los estudios individuales

La evaluación del riesgo de sesgo es un paso crucial en las revisiones sistemáticas para determinar la validez interna de los estudios incluidos. Dado que la presente revisión incluyó estudios con diversos diseños metodológicos (estudios empíricos, revisiones de literatura, análisis de casos, etc.), no se empleó una herramienta estandarizada de evaluación de riesgo de sesgo. En cambio, se realizó una evaluación crítica de la calidad de la evidencia de cada estudio, considerando la adecuación del diseño del estudio al objetivo de la investigación, la rigurosidad en la recolección y análisis de datos, y la transparencia en la presentación de resultados y limitaciones. Se examinaron aspectos como la validez interna, la validez externa (generalización de los resultados), la precisión de las mediciones y el control de posibles factores de confusión (Petticrew & Roberts, 2006).

Métodos de síntesis

Para analizar e integrar los hallazgos de los estudios incluidos, se empleó una síntesis narrativa de la evidencia. Este enfoque cualitativo permitió describir, analizar y sintetizar los resultados de los estudios, identificando patrones, similitudes, diferencias y contradicciones en las evidencias (Popay et al., 2006). La síntesis narrativa se estructuró en torno a los temas clave emergentes de la literatura, como los dilemas éticos más comunes, las implicaciones sociales más relevantes, y las principales recomendaciones para la implementación ética de la IA en la administración empresarial.

RESULTADOS

Esta sección presenta los resultados del proceso de revisión sistemática, siguiendo las directrices PRISMA. Se describen los resultados de la búsqueda, las características de los estudios incluidos, una síntesis de los hallazgos y una evaluación de la certeza de la evidencia.

Características de los Estudios:

Los estudios incluidos en la revisión, publicados entre 2018 y 2024, abarcan una variedad de diseños metodológicos, desde análisis conceptuales y revisiones de literatura hasta estudios de caso y propuestas de marcos teóricos. Esta diversidad metodológica enriquece la revisión al proporcionar una visión amplia y multifacética de las implicaciones éticas y sociales de la IA en la administración. La Tabla 2-4 presenta un resumen completo de las características de cada estudio, incluyendo el autor, el año de publicación, el título, el contexto específico de la investigación, el diseño del estudio, los principales dilemas éticos abordados, las implicaciones sociales identificadas y las directrices o recomendaciones propuestas.

Tabla 2-3

Características principales de estudios sobre ética e IA

Autor(es)	Dilemas éticos	Implicaciones sociales	Directrices propuestas
Rosemann & Zhang	Confiabilidad, privacidad, impacto en el empleo	Equidad en el acceso, justicia	RRI, estándares internacionales
Harlow	Privacidad, uso de datos, sesgo algorítmico	Discriminación, desigualdad	Código de ética, gestión responsable de datos
Santoni de Sio & Mecacci	Brechas de responsabilidad (culpabilidad, rendición de cuentas)	Confianza, justicia, control humano	Diseño de sistemas con “control humano significativo”
Brendel et al.	Toma de decisiones, consideraciones éticas	Impacto en los stakeholders	Marco EMMA (Gestión Ética de la IA)
Djeffal	Gobernanza, regulación, transparencia	Confianza, justicia, rendición de cuentas	Directrices éticas, regulación moderada

Autor(es)	Dilemas éticos	Implicaciones sociales	Directrices propuestas
Weber-Lewerenz	Responsabilidad digital corporativa, ética en el diseño	Sostenibilidad, impacto social de la tecnología	Directrices CDR, ética en el diseño
Elliott et al.	Equidad, transparencia, algoritmos opacos	Sociedad digital equitativa, inclusión	CDR como mecanismo de gobernanza, transparencia
Benefo et al.	Diversas implicaciones ELSE, necesidad de un enfoque multidisciplinario	Impacto en diferentes niveles (individual, social, global)	Enfoque multidisciplinario, participación de stakeholders
Carter	Regulación, gobernanza, rol de los profesionales de la información	Confianza, mitigación de riesgos	Marcos regulatorios, participación activa de profesionales de la información
Zaman	Control, responsabilidad, toma de decisiones	Autonomía humana, impacto social a largo plazo	Necesidad de reglas fundacionales, equilibrio entre control y libertad
Carrillo	Relación entre ética y derecho en la IA, gobernanza	Implicaciones legales del uso de la IA en la sociedad	Derecho internacional como marco regulatorio
Kerr, Barry & Kelleher	Expectativas sociales vs. realidad de la “IA ética”, gobernanza	Brecha entre expectativas y realidad, necesidad de regulación	Herramientas, prácticas y marcos regulatorios para una IA ética
Kieslich, Keller & Starke	Principios éticos del diseño de la IA, preferencias sociales	Importancia de la percepción pública en el diseño ético de la IA	Diseño centrado en el ser humano, considerar diversidad de valores
Nikadimovs & Vēvere	Responsabilidad social universitaria en el uso de IA	Percepción de <i>stakeholders</i> sobre IA y responsabilidad social	Necesidad de evaluar y considerar la percepción de <i>stakeholders</i>
Ali et al.	No especificados en el texto proporcionado		

Nota: Elaboración propia a partir de la revisión de la literatura.

Análisis bibliométrico:

Un análisis del año de publicación de los estudios incluidos revela un interés creciente en el tema de la ética de la IA en la administración empresarial. Mientras que solo se identificó un estudio publicado en 2018, el número aumentó a dos en 2020, cuatro en 2021, tres en 2022, uno en 2023 y cuatro en 2024 (hasta la fecha de la última búsqueda). Este crecimiento en la producción científica sugiere que la ética de la IA se está convirtiendo en un tema cada vez más relevante para los investigadores y profesionales de la administración.

El análisis de la autoría y afiliación institucional de los estudios muestra una diversidad de autores e instituciones de diferentes países, lo que indica un interés global en el tema. Sin embargo, no se identificaron autores o instituciones que dominen claramente la producción científica en este campo específico. Este hallazgo sugiere que la ética de la IA en la administración es un campo de investigación relativamente nuevo y en expansión, con oportunidades para una mayor colaboración y desarrollo de una comunidad académica más consolidada.

Análisis de contenido:

Un análisis preliminar del contenido de los estudios, basado en la frecuencia de palabras clave y la co-ocurrencia de términos, revela que los siguientes temas son centrales en el debate sobre la ética de la IA en la administración:

- **Responsabilidad:** La cuestión de quién es responsable de las decisiones y acciones tomadas por sistemas de IA en el contexto empresarial es un tema recurrente en la literatura.
- **Transparencia:** La necesidad de sistemas de IA transparentes y explicables, que permitan a los usuarios comprender cómo se toman las decisiones, se destaca como un factor crítico para generar confianza.
- **Sesgo y discriminación:** El riesgo de sesgo algorítmico y sus posibles consecuencias discriminatorias en la toma de decisiones empresariales se discuten ampliamente.
- **Privacidad y protección de datos:** La recopilación y el uso de grandes conjuntos de datos por parte de los sistemas de IA plantean preocupaciones sobre la privacidad individual y la seguridad de la información.

- **Impacto en el empleo:** Las implicaciones de la automatización impulsada por la IA en el mercado laboral, incluyendo la posible pérdida de empleos y la necesidad de nuevas habilidades, son objeto de debate.

Riesgo de sesgo de los estudios individuales

La evaluación del riesgo de sesgo en cada estudio individual se centró en identificar posibles factores que pudieran haber afectado la validez interna o la generalización de los hallazgos. Se consideró que la mayoría de los estudios presentaban al menos un tipo de limitación metodológica, lo que es común en la investigación en ciencias sociales, especialmente en un campo emergente como la ética de la IA. Algunas de las limitaciones identificadas fueron:

- **Tamaños de muestra pequeños:** Algunos estudios se basaron en muestras pequeñas o en estudios de caso únicos, lo que limita la generalización de los resultados a otros contextos.
- **Falta de grupos de control:** Algunos estudios no utilizaron grupos de control, lo que dificulta determinar si los efectos observados se deben realmente a la variable de interés (la IA) o a otros factores.
- **Sesgo de selección:** En algunos casos, la selección de participantes o casos de estudio podría haber introducido sesgos que afectaron los resultados.
- **Falta de longitudinalidad:** La mayoría de los estudios fueron transversales, es decir, recolectaron datos en un solo punto en el tiempo. Los estudios longitudinales, que recopilan datos a lo largo del tiempo, son más adecuados para evaluar el impacto a largo plazo de la IA.

A pesar de estas limitaciones, se consideró que la calidad general de la evidencia era suficiente para extraer conclusiones significativas, ya que los estudios seleccionados para la revisión provenían de fuentes académicas de renombre y habían sido sometidos a un proceso de revisión por pares.

Resultados de los estudios individuales

Cada uno de los 15 estudios incluidos proporcionó información única y valiosa sobre los desafíos éticos y sociales de la IA en la administración empresarial. Algunos estudios, como el de Harlow (2018), adoptaron un enfoque amplio al analizar una gama de dilemas éticos relacionados con la privacidad, el sesgo algorítmico y la necesidad de una gestión responsable de los datos. Otros estudios se centraron en

aspectos específicos, como la investigación de Santoni de Sio & Mecacci (2021), que profundizó en las “brechas de responsabilidad” que surgen con la IA.

Es importante destacar que muchos estudios resaltaron la necesidad de un enfoque proactivo para abordar los desafíos éticos de la IA. Por ejemplo, Brendel et al. (2021) propusieron el marco EMMA (Ethical Management of AI) para ayudar a las empresas a integrar consideraciones éticas en la gestión de la IA. Asimismo, Djeflal (2020) abogó por la creación de directrices éticas y marcos regulatorios claros para guiar la implementación responsable de la IA en las organizaciones.

Resultados de la síntesis

La síntesis narrativa de la evidencia reveló una serie de temas clave que emergen de la literatura sobre ética de la IA en la administración empresarial:

- **Preocupaciones sobre la equidad y la discriminación:** La capacidad de la IA para procesar grandes cantidades de datos y tomar decisiones automatizadas plantea preocupaciones sobre la amplificación de sesgos existentes en los datos, lo que podría llevar a prácticas discriminatorias en áreas como la contratación, la promoción y la asignación de recursos (Rosemann & Zhang, 2022).
- **Necesidad de transparencia y explicabilidad:** La falta de transparencia en los procesos de toma de decisiones de los sistemas de IA, especialmente en aquellos basados en algoritmos complejos de aprendizaje automático, se identificó como un obstáculo importante para garantizar la responsabilidad y generar confianza (Elliott et al., 2021). La explicabilidad, es decir, la capacidad de comprender y explicar las razones detrás de las decisiones de la IA, se considera esencial para identificar posibles sesgos, corregir errores y garantizar que las decisiones sean justas e imparciales.
- **Importancia de la responsabilidad y la gobernanza:** Los estudios coincidieron en la necesidad de establecer mecanismos claros de responsabilidad para las decisiones tomadas por o con la ayuda de sistemas de IA (Djeflal, 2020). Esto implica definir roles y responsabilidades dentro de las organizaciones, implementar mecanismos de auditoría y supervisión, y establecer vías de recurso para abordar posibles daños o resultados injustos. La gobernanza de la IA, que abarca los procesos, políticas y estructuras organizacionales que guían el desarrollo y la implementación de la IA, se considera crucial para garantizar que la IA se utilice de manera ética y responsable.

- **Impacto en el empleo y las habilidades:** Si bien la IA puede generar nuevas oportunidades de empleo y aumentar la productividad, también existe preocupación por la posible desplazamiento de trabajadores en ciertas ocupaciones (Carter, 2020). La naturaleza cambiante del trabajo en la era de la IA requiere que las empresas inviertan en la capacitación y el desarrollo de habilidades de sus empleados para que puedan adaptarse a las nuevas demandas del mercado laboral.
- **Implicaciones para la privacidad y la protección de datos:** La capacidad de la IA para recopilar, almacenar y analizar grandes cantidades de datos, incluyendo datos personales, plantea desafíos importantes para la privacidad y la protección de datos (Harlow, 2018). Las empresas deben garantizar que el uso de la IA cumpla con las regulaciones de protección de datos, como el Reglamento General de Protección de Datos (RGPD), y adoptar un enfoque centrado en la privacidad en el diseño e implementación de sistemas de IA.

Certeza de la evidencia

La certeza de la evidencia se evaluó utilizando el enfoque GRADE (Grading of Recommendations Assessment, Development and Evaluation), ampliamente utilizado en revisiones sistemáticas para calificar la calidad de la evidencia y la fuerza de las recomendaciones (Guyatt et al., 2008). En general, la certeza de la evidencia para los temas clave identificados en esta revisión se consideró moderada. Esto significa que existe la posibilidad de que futuras investigaciones, con diseños metodológicos más robustos y tamaños de muestra más grandes, puedan cambiar las estimaciones del efecto o la dirección de los hallazgos.

Factores que limitan la certeza de la evidencia:

- **Naturaleza emergente del campo:** La ética de la IA en la administración es un campo relativamente nuevo y en rápida evolución, y la base de evidencia aún está en desarrollo.
- **Limitaciones metodológicas:** Muchos estudios presentaban limitaciones metodológicas, como se describió anteriormente, lo que limita la certeza de los hallazgos.
- **Heterogeneidad de los estudios:** Los estudios incluidos en la revisión variaron en términos de diseños de investigación, poblaciones estudiadas, contextos y resultados medidos, lo que dificulta la comparación y generalización de los resultados.

DISCUSIÓN

Las implicaciones de la inteligencia artificial (IA) en el área profesional de la administración son profundas y multifacéticas. Esta revisión sistemática ha revelado un creciente interés por comprender y abordar los desafíos éticos y sociales que surgen con la adopción de la IA en la gestión empresarial. Los administradores se enfrentan a la necesidad de integrar la IA de manera responsable, considerando no solo sus beneficios potenciales en eficiencia y productividad, sino también sus posibles impactos en la equidad, la transparencia, la responsabilidad y el bienestar de los trabajadores. Como señalan Fountaine et al. (2021), “La IA no es simplemente una herramienta más; es una fuerza transformadora que redefine los fundamentos de la gestión empresarial”.

Un tema central que emerge de los estudios revisados es la preocupación por la equidad en la toma de decisiones automatizada. Si bien los algoritmos de IA ofrecen la promesa de decisiones más objetivas y eficientes, también pueden perpetuar y amplificar los sesgos presentes en los datos utilizados para su entrenamiento (Rosemann & Zhang, 2022). Esto puede llevar a resultados discriminatorios, especialmente en áreas sensibles como la contratación, la promoción y la asignación de recursos. Un estudio reciente de Obermeyer et al. (2019) encontró que un algoritmo ampliamente utilizado en el sector de la salud para asignar atención médica a pacientes presentaba un sesgo racial significativo, perjudicando a los pacientes afroamericanos.

La revisión también destaca la importancia de una gestión ética de los datos. Harlow (2018) advierte que los datos utilizados para entrenar los algoritmos de IA pueden reflejar prejuicios y sesgos existentes en la sociedad, lo que puede conducir a resultados discriminatorios. Para garantizar la equidad y la justicia en la toma de decisiones automatizada, las empresas deben implementar estrategias para identificar y mitigar el sesgo en los datos, como la auditoría de algoritmos, la recopilación de datos más representativos y el desarrollo de algoritmos que sean explícitamente diseñados para promover la equidad (Holstein et al., 2019).

La transparencia y la explicabilidad se identifican como elementos cruciales para una implementación responsable de la IA en la administración. La “caja negra” que envuelve a muchos sistemas de IA genera desconfianza y dificulta la identificación y corrección de posibles sesgos o errores (Elliott et al., 2021). Un estudio de Mittelstadt et al. (2019) sugiere que la explicabilidad de la IA es esencial para facilitar la auditoría, permitir la contestación de decisiones, mejorar la confiabilidad y fomentar la innovación responsable.

La cuestión de la responsabilidad por las decisiones y acciones tomadas por sistemas de IA es un tema complejo y en evolución. Santoni de Sio & Mecacci (2021) argumentan que las “brechas de responsabilidad” que surgen con la IA, como la dificultad para atribuir culpabilidad o garantizar la rendición de cuentas, requieren un replanteamiento de los marcos éticos y legales tradicionales. Djeflal (2020) y Carter (2020) coinciden en la necesidad de establecer directrices éticas y marcos regulatorios claros para guiar la implementación responsable de la IA, pero el debate sobre el grado de regulación necesario para equilibrar la innovación con la protección de los derechos sigue abierto.

Las implicaciones de la IA en el empleo y las habilidades requieren una gestión proactiva por parte de las empresas. Si bien la IA puede crear nuevas oportunidades y aumentar la productividad, también existe la preocupación de que la automatización pueda llevar a la pérdida de empleos en ciertas áreas. Como señalan Brynjolfsson & McAfee (2014), “La IA no solo está automatizando tareas individuales, sino que está comenzando a automatizar procesos empresariales completos”.

Las empresas deben invertir en programas de capacitación y desarrollo de habilidades para ayudar a sus empleados a adaptarse a las nuevas demandas del mercado laboral (Brendel et al., 2021; Weber-Lewerenz, 2021). Además, las empresas deben considerar las implicaciones éticas de la automatización, buscando un equilibrio entre la eficiencia y el bienestar de los trabajadores.

La gestión de la privacidad y la protección de datos es esencial para construir confianza en la era de la IA. Harlow (2018) enfatiza la necesidad de políticas de privacidad sólidas que protejan los datos personales de los empleados y clientes. Floridi (2019) argumenta que la ética de la información debe ser un principio guía en el desarrollo e implementación de sistemas de IA, garantizando que los datos se utilicen de manera responsable y respetuosa.

Implicaciones para la gestión de recursos humanos:

Las implicaciones de la IA para la gestión de recursos humanos (GRH) son profundas y requieren un replanteamiento de los enfoques tradicionales. La IA no solo está automatizando tareas administrativas dentro de la GRH, sino que también está transformando la forma en que las empresas reclutan, seleccionan, capacitan, evalúan y gestionan a sus empleados.

Los administradores de recursos humanos deben adquirir una comprensión profunda de los desafíos éticos y sociales de la IA para poder integrarla de manera responsable en sus prácticas. Algunas de las implicaciones clave para la GRH incluyen:

- **Redefinición de roles y habilidades:** La automatización de tareas administrativas liberará a los profesionales de la GRH para centrarse en actividades más estratégicas y de valor agregado, como el desarrollo de talento, la gestión del cambio y la creación de una cultura organizacional positiva (Davenport & Kirby, 2016).
- **Ética en la contratación y promoción:** La IA puede ayudar a eliminar los sesgos humanos en los procesos de contratación y promoción, pero también presenta el riesgo de perpetuar los sesgos presentes en los datos (Obermeyer et al., 2019). Se necesitan mecanismos de auditoría y control para garantizar la equidad y la transparencia en estos procesos.
- **Personalización del desarrollo de talento:** La IA puede utilizarse para crear programas de capacitación y desarrollo más personalizados, adaptando el contenido y el ritmo de aprendizaje a las necesidades individuales de los empleados (Tursunbayeva, 2018).
- **Gestión de la diversidad e inclusión:** La IA puede ayudar a identificar y abordar las barreras a la diversidad y la inclusión en el lugar de trabajo, pero también puede reforzar las desigualdades existentes si no se implementa de manera cuidadosa (Tambe et al., 2019).

Este tema es crucial para los administradores de recursos humanos que buscan crear entornos de trabajo atractivos y sostenibles, y minimizar los costos asociados con la alta rotación de personal.

CONCLUSIONES

Esta revisión sistemática ha examinado los dilemas éticos y las implicaciones de la responsabilidad social que surgen con la adopción de la IA en la administración de empresas. Los hallazgos confirman los objetivos de la investigación, mostrando que la integración de la IA en la gestión empresarial presenta una serie de desafíos éticos, que incluyen la equidad, la transparencia, la responsabilidad, el impacto en el empleo y la privacidad de los datos.

La revisión también ha identificado la necesidad de desarrollar e implementar directrices éticas claras y marcos de gobernanza robustos para guiar la implementación responsable de la IA en las empresas. Las conclusiones de esta revisión tienen im-

plicaciones prácticas para los profesionales de la administración, especialmente para aquellos involucrados en la gestión de recursos humanos. Se necesitan estrategias proactivas para abordar los desafíos éticos, crear una cultura ética sólida y garantizar que la IA se utilice para promover un entorno de trabajo justo, transparente y sostenible.

REFERENCIAS BIBLIOGRÁFICAS

- Ali, K., Alzaidi, M., Al-Fraihat, D., & Elamir, A. M. (2023). Artificial intelligence: benefits, application, ethical issues, and organizational responses. En *Intelligent Sustainable Systems: selected Papers of World S4 2022, volume 1* (pp. 685-702). Springer Nature Singapore. https://link.springer.com/chapter/10.1007/978-981-19-7660-5_62
- Benefo, E. O., Tingler, A., White, M., Cover, J., Torres, L., Broussard, C., ... & Patra, D. (2022). Ethical, legal, social, and economic (ELSE) implications of artificial intelligence at a global level: a scientometrics approach. *AI and Ethics*, 2(4), 667-682. <https://link.springer.com/article/10.1007/s43681-021-00124-6>
- Brendel, A. B., Mirbabaie, M., Lembcke, T. B., & Hofeditz, L. (2021). Ethical management of artificial intelligence. *Sustainability*, 13(4), 1974. <https://www.mdpi.com/2071-1050/13/4/1974>
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company. [https://books.google.es/books?hl=es&lr=&id=WiKwAgAAQBAJ&oi=fnd&pg=PA1&dq=Brynjolfsson,+E.,+%26+McAfee,+A.+\(2014\).+The+second+machine+age:+Work,+progress,+and+prosperity+in+a+time+of+brilliant+technologies.+W.+W.+Norton+%26+Company.&ots=4-VoWh-tgi&sig=gQOxZmtQ_K7o8bfRF92Ou3tolSQ](https://books.google.es/books?hl=es&lr=&id=WiKwAgAAQBAJ&oi=fnd&pg=PA1&dq=Brynjolfsson,+E.,+%26+McAfee,+A.+(2014).+The+second+machine+age:+Work,+progress,+and+prosperity+in+a+time+of+brilliant+technologies.+W.+W.+Norton+%26+Company.&ots=4-VoWh-tgi&sig=gQOxZmtQ_K7o8bfRF92Ou3tolSQ)
- Carrillo, M. R. (2020). Artificial intelligence: From ethics to law. *Telecommunications Policy*, 44(6), 101937. <https://www.sciencedirect.com/science/article/pii/S030859612030029X>
- Carter, D. (2020). Regulation and ethics in artificial intelligence and machine learning technologies: Where are we now? Who is responsible? Can the information professional play a role? *Business Information Review*, 37(2), 60-68. <https://journals.sagepub.com/doi/abs/10.1177/0266382120923962>

- Davenport, T. H., & Kirby, J. (2016). *Only humans need apply: Winners and losers in the age of smart machines*. Harper Business. https://leadersexcellence.com/wp-content/uploads/dlm_uploads/2016/08/Davenport-Leaders-Excellence-presentation.pdf
- Djeffal, C. (2020). Artificial intelligence and public governance: normative guidelines for artificial intelligence in government and public administration. *Regulating Artificial Intelligence*, 277-293. https://link.springer.com/chapter/10.1007/978-3-030-32361-5_12
- Elliott, K., Price, R., Shaw, P., Spiliotopoulos, T., Ng, M., Coopamootoo, K., & Van Moorsel, A. (2021). Towards an equitable digital society: artificial intelligence (AI) and corporate digital responsibility (CDR). *Society*, 58(3), 179-188. <https://link.springer.com/article/10.1007/s12115-021-00594-8>
- Floridi, L. (2019). The ethics of information. En F. van der Hoek (Ed.), *The Oxford Handbook of Philosophy of Information* (pp. 443-462). Oxford University Press. [https://books.google.es/books?hl=es&lr=&id=_XHcAAAAQBAJ&oi=fnd&pg=PP1&dq=Floridi,+L.+\(2019\).+The+ethics+of+information.+En+F.+van+der+Hoek+\(Ed.\),+The+Oxford+Handbook+of+Philosophy+of+Information+\(pp.+443-462\).+Oxford+University+Press.&ots=f_hD5_RvWX&sig=-VGvY9MjV_G54RXJLEfqxQA3N8ZI](https://books.google.es/books?hl=es&lr=&id=_XHcAAAAQBAJ&oi=fnd&pg=PP1&dq=Floridi,+L.+(2019).+The+ethics+of+information.+En+F.+van+der+Hoek+(Ed.),+The+Oxford+Handbook+of+Philosophy+of+Information+(pp.+443-462).+Oxford+University+Press.&ots=f_hD5_RvWX&sig=-VGvY9MjV_G54RXJLEfqxQA3N8ZI)
- Fountaine, T., McCarthy, B., & Saleh, T. (2021). Building the AI-powered organization. *Harvard Business Review*, 99(4), 62-73. https://wuyuansheng.com/doc/Databricks-AI-Powered-Org__Article-Licensing-July21-1.pdf
- Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *BMJ*, 331(7524), 1065-1069. <https://www.bmj.com/content/331/7524/1064.short>
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., ... & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650), 924-926. <https://www.bmj.com/content/336/7650/924.short>
- Harlow, H. (2018, Septiembre). Ethical concerns of artificial intelligence, big data and data analytics. En *European Conference on Knowledge Management* (pp. 316-323). Academic Conferences International Limited. <https://search.proquest>

[com/openview/56146799514cd6f3c4e92e82079fb128/1?pq-origsite=gscholar&cbl=1796412](https://openview/56146799514cd6f3c4e92e82079fb128/1?pq-origsite=gscholar&cbl=1796412)

- Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? En *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). <https://dl.acm.org/doi/abs/10.1145/3290605.3300830>
- Kerr, A., Barry, M., & Kelleher, J. D. (2020). Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance. *Big Data & Society*, 7(1), 2053951720915939. <https://journals.sagepub.com/doi/abs/10.1177/2053951720915939>
- Kieslich, K., Keller, B., & Starke, C. (2022). Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence. *Big Data & Society*, 9(1), 20539517221092956. <https://journals.sagepub.com/doi/abs/10.1177/20539517221092956>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2019). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 6(1), 2053951719839315. <https://journals.sagepub.com/doi/abs/10.1177/2053951716679679>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://www.acpjournals.org/doi/abs/10.7326/0003-4819-151-4-200908180-00135>
- Nikadimovs, O., & Vēvere, V. (2024, junio). The Use of Generative Artificial Intelligence in Higher Education: University Social Responsibility and Stakeholders' perceptions. En *Environment. Technologies. Resources. Proceedings of the International Scientific and Practical Conference* (Vol. 2, pp. 226-231). <https://journals.ru.lv/index.php/ETR/article/view/8015>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453. <https://www.science.org/doi/abs/10.1126/science.aax2342>
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing. https://books.google.es/books?hl=es&lr=&id=ZwZ1_xU3E80C&oi=fnd&pg=PR5&dq=Petticrew,+M.,+%26+Ro-

- berts,+H.+(2006).+Systematic+reviews+in+the+social+sciences:+A+practical+guide.+Blackwell+Publishing.&ots=w_Q_xQIUlt&sig=wzaboMWCLs-Vs4M4w3suRp70mdpM
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme*. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ed8b23836338f6fdea0cc55e161b0fc5805f9e27>
- Rosemann, A., & Zhang, X. (2022). Exploring the social, ethical, legal, and responsibility dimensions of artificial intelligence for health-a new column in Intelligent Medicine. *Intelligent Medicine*, 2(02), 103-109. <https://mednexus.org/doi/abs/10.1016/j.imed.2021.12.002>
- Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, 34(4), 1057-1084. <https://link.springer.com/article/10.1007/s13347-021-00450-x>
- Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15-42. <https://journals.sagepub.com/doi/abs/10.1177/0008125619867910>
- Tursunbayeva, A. (2019). Human resource technology disruptions and their implications for human resources management in healthcare organizations. *BMC health services research*, 19(1), 268. <https://link.springer.com/article/10.1186/s12913-019-4068-3>
- Weber-Lewerenz, B. (2021). Corporate digital responsibility (CDR) in construction engineering—ethical guidelines for the application of digital transformation and artificial intelligence (AI) in user practice. *SN Applied Sciences*, 3, 1-25. <https://link.springer.com/article/10.1007/s42452-021-04776-1>
- Zaman, B. U. (2024). The weighty responsibility of creating ai navigating control and ethics. <https://www.preprints.org/manuscript/202404.1900>