

ARTÍCULO

Gobierno de la inteligencia artificial y su integración con los principios ESG en corporaciones latinoamericanas

Governança da inteligência artificial e sua integração com os princípios ESG em corporações latino-americanas

Artificial intelligence governance and its integration with ESG principles in Latin American corporations

Lincon Lopes

¹ Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), São Paulo, SP, Brasil. ORCID: <https://orcid.org/0000-0002-6281-1226>

ENVIADO	ACEPTADO	PUBLICADO	DOI
17/03/2025	25/10/2025	05/12/2025	en proceso

RESUMEN

La adopción corporativa de la inteligencia artificial plantea dilemas éticos y de gobernanza que afectan de forma directa el desempeño ambiental, social y de gobernanza (ESG) de las organizaciones. Este trabajo examina cómo las empresas de América Latina incorporan, o dejan de incorporar, marcos de gobernanza algorítmica capaces de garantizar transparencia, mitigación de sesgos y rendición de cuentas. Mediante una revisión teórica integrativa de la literatura sobre gobernanza corporativa, ética de la inteligencia artificial y sostenibilidad empresarial, se construye un modelo conceptual que articula tres capacidades organizacionales: supervisión algorítmica, explicabilidad técnica y participación de las partes interesadas. El argumento central sostiene que la inteligencia artificial solo genera valor sostenible cuando queda subordinada a directrices éticas explícitas y a una supervisión institucional independiente del área que desarrolla o adquiere la tecnología. Se discuten además las particularidades del contexto latinoamericano, marcado por marcos regulatorios heterogéneos, estructuras de propiedad

concentradas y una desigualdad estructural que amplifica los riesgos de exclusión algorítmica. El trabajo concluye con una agenda de investigación empírica y con recomendaciones prácticas para consejos de administración y comités de sostenibilidad.

Palabras clave: gobernanza de IA, ESG, ética organizacional, creación de valor, sostenibilidad corporativa

RESUMO

A adoção corporativa da inteligência artificial apresenta dilemas éticos e de governança que afetam diretamente o desempenho ambiental, social e de governança (ESG) das organizações. Este trabalho examina como as empresas da América Latina incorporam, ou deixam de incorporar, marcos de governança algorítmica capazes de garantir transparência, mitigação de vieses e prestação de contas. Por meio de uma revisão teórica integrativa da literatura sobre governança corporativa, ética da inteligência artificial e sustentabilidade empresarial, constrói-se um modelo conceitual que articula três capacidades organizacionais: supervisão algorítmica, explicabilidade técnica e participação das partes interessadas. O argumento central sustenta que a inteligência artificial só gera valor sustentável quando subordinada a diretrizes éticas explícitas e a uma supervisão institucional independente da área que desenvolve ou adquire a tecnologia. Discutem-se ainda as particularidades do contexto latino-americano, marcado por marcos regulatórios heterogêneos, estruturas de propriedade concentradas e uma desigualdade estrutural que amplifica os riscos de exclusão algorítmica. O trabalho conclui com uma agenda de pesquisa empírica e recomendações práticas para conselhos de administração e comitês de sustentabilidade.

Palavras-chave: governança de IA, ESG, ética organizacional, criação de valor, sustentabilidade corporativa

ABSTRACT

The corporate adoption of artificial intelligence raises ethical and governance dilemmas that directly affect the environmental, social, and governance (ESG) performance of organizations. This paper examines how companies in Latin America incorporate, or fail to incorporate, algorithmic governance frameworks capable of ensuring transparency, bias mitigation, and accountability. Through an integrative theoretical review of the literature on corporate governance, artificial intelligence ethics, and

corporate sustainability, a conceptual model is built around three organizational capabilities: algorithmic oversight, technical explainability, and stakeholder engagement. The central argument holds that artificial intelligence only generates sustainable value when it is subordinated to explicit ethical guidelines and to institutional oversight independent from the unit that develops or acquires the technology. The paper also discusses the particularities of the Latin American context, marked by heterogeneous regulatory frameworks, concentrated ownership structures, and structural inequality that amplifies the risks of algorithmic exclusion. The study closes with an agenda for empirical research and practical recommendations for boards of directors and sustainability committees.

Keywords: AI governance, ESG, organizational ethics, value creation, corporate sustainability

1 Introducción

La incorporación acelerada de tecnologías basadas en inteligencia artificial al núcleo de las operaciones corporativas ha cambiado los términos de la estrategia competitiva. En un panorama económico marcado por la hiperconectividad digital y por mercados cada vez más volátiles, la transformación digital tiende a presentarse casi como una condición de supervivencia para cualquier empresa que aspire a seguir siendo relevante. Sin embargo, este avance tecnológico ocurre en paralelo a una exigencia institucional, social y regulatoria creciente en torno a la sostenibilidad corporativa, sintetizada en los criterios ambientales, sociales y de gobernanza conocidos por la sigla ESG. Si bien la adopción de la inteligencia artificial suele justificarse por su capacidad de mejorar la eficiencia operativa y la rentabilidad financiera, su efecto sobre las dimensiones ESG de las organizaciones es ambivalente y, en no pocos casos, fuente de riesgos sistémicos que los órganos de dirección ejecutiva no siempre evalúan con el rigor necesario.

El problema central abordado en este estudio es la desconexión, tanto operativa como institucional, que suele observarse entre las estrategias de transformación digital basadas en inteligencia artificial y los marcos de gobernanza y responsabilidad social corporativa. En América Latina, las corporaciones enfrentan el desafío de adoptar tecnologías avanzadas en entornos institucionales caracterizados por marcos regulatorios fragmentados y por una alta sensibilidad social ante la desigualdad estructural que afecta a la región desde hace décadas (La Porta, López-de-Silanes, Shleifer y Vishny, 2000). La ausencia de un modelo integrado de

gobernanza de la inteligencia artificial y de principios ESG expone a las empresas a fallas éticas severas, a la pérdida de confianza de inversionistas institucionales y a daños reputacionales que pueden comprometer su continuidad operativa en el mediano plazo.

Conviene precisar, desde el inicio, qué se entiende aquí por gobernanza de la inteligencia artificial. Siguiendo a Wirtz, Weyerer y Geyer (2019) y a Cihon, Maas y Kemp (2020), se entiende como el conjunto de estructuras, procesos, normas y mecanismos de control que una organización establece para dirigir, supervisar y auditar el diseño, la adquisición y el uso de sistemas algorítmicos, de modo que sus decisiones sean explicables, auditables y compatibles con los valores declarados de la empresa. Esta definición se distingue de la simple gestión de riesgos tecnológicos porque incorpora una dimensión normativa: no basta con que un sistema funcione de manera eficiente, también debe poder justificarse ante las partes afectadas por sus decisiones.

La literatura sobre administración estratégica y economía institucional ha señalado, desde hace tiempo, que la transformación digital no es un proceso neutro ni desprovisto de valores. Por el contrario, reconfigura las relaciones de poder dentro de la organización, altera las estructuras de trabajo y modifica la distribución de externalidades negativas sobre las comunidades locales y los ecosistemas (Zuboff, 2019). Las corporaciones que implementan sistemas algorítmicos sin salvaguardas éticas suficientes tienden a priorizar ganancias de eficiencia de corto plazo en detrimento de la equidad social y de la integridad institucional, lo que erosiona, con el tiempo, los cimientos de la sostenibilidad corporativa.

Este trabajo se apoya en las contribuciones de Agrawal, Gans y Goldfarb (2018) sobre la economía de la predicción, en los marcos de gobernanza sociotécnica propuestos por Gasser y Almeida (2017), y en la literatura sobre ética aplicada a sistemas algorítmicos desarrollada por Mittelstadt, Allo, Taddeo, Wachter y Floridi (2016) y por Jobin, Ienca y Vayena (2019). Se sostiene que la gobernanza algorítmica no puede limitarse al cumplimiento normativo superficial; necesita constituirse como una capacidad dinámica capaz de reconfigurar las rutinas organizacionales frente a las demandas cambiantes de las partes interesadas (Teece, 2007). Esta perspectiva coincide con la advertencia de Whittlestone, Nyrup, Alexandrova y Cave (2019), quienes señalan que los principios éticos declarativos, sin mecanismos concretos de implementación, tienden a permanecer en el plano retórico y a no modificar las prácticas organizacionales reales.

La pregunta que orienta esta investigación es la siguiente: ¿cómo pueden las corporaciones latinoamericanas estructurar un marco de gobernanza de la inteligencia artificial que asegure su alineación efectiva con los principios ESG y que, al mismo tiempo, promueva la creación de valor sostenible? El objetivo principal del artículo es proponer un modelo teórico y normativo que articule la supervisión algorítmica con las tres dimensiones de la sostenibilidad corporativa. De manera más específica, el trabajo busca: a) sistematizar los principales marcos conceptuales sobre gobernanza de inteligencia artificial y su relación con la teoría de la agencia y la teoría de las partes interesadas; b) identificar los puntos de fricción y de sinergia entre los objetivos de la transformación digital y los estándares ESG reconocidos internacionalmente; y c) formular recomendaciones prácticas para consejos de administración, comités de auditoría y áreas de sostenibilidad en organizaciones que operan en la región.

La pertinencia de este análisis para el contexto latinoamericano no es menor. A diferencia de economías con marcos regulatorios de protección de datos consolidados, como la Unión Europea tras la entrada en vigor del Reglamento General de Protección de Datos, la mayoría de los países de la región cuenta con legislaciones incipientes o fragmentadas en materia de inteligencia artificial (Aguerre, 2021). Esta situación traslada buena parte de la responsabilidad regulatoria a la autorregulación corporativa, lo que hace aún más urgente comprender cómo los consejos de administración y la alta dirección pueden institucionalizar mecanismos de control que suplan, al menos parcialmente, los vacíos normativos existentes.

El texto está organizado en siete secciones. La segunda desarrolla el marco teórico sobre gobernanza algorítmica, teoría de la agencia ampliada y criterios ESG. La tercera describe el enfoque metodológico integrativo empleado. La cuarta examina las intersecciones críticas entre la inteligencia artificial y los tres pilares ESG, con énfasis en la evidencia disponible para la región. La quinta aborda los mecanismos institucionales concretos para la sostenibilidad algorítmica. La sexta analiza las implicaciones gerenciales para la alta dirección y los consejos de administración. Por último, la séptima sección presenta las consideraciones finales, las limitaciones del estudio y una agenda de investigación futura.

2 Marco teórico: gobernanza de la inteligencia artificial y criterios ESG

El análisis de la gobernanza algorítmica en el contexto corporativo latinoamericano revela varias capas de complejidad institucional que conviene desagregar antes de proponer cualquier modelo integrador. Por un

lado, la presión competitiva global empuja a las empresas a adoptar soluciones de inteligencia artificial para optimizar cadenas de valor y reducir costos de transacción. Por otro lado, la fragilidad histórica de los marcos regulatorios locales en materia de protección de datos personales, equidad algorítmica y derechos laborales digitales deja a trabajadores y consumidores en una posición de vulnerabilidad frente a decisiones cada vez más automatizadas. Esta asimetría regulatoria aumenta los riesgos reputacionales y operativos para las corporaciones de la región y hace necesaria una autorregulación ética seria, respaldada por la adopción voluntaria de estándares internacionales de debida diligencia.

La literatura de gestión estratégica ha documentado que la implementación de tecnologías disruptivas altera los equilibrios de poder dentro de las organizaciones. Los departamentos de tecnología y ciencia de datos suelen adquirir un protagonismo considerable, y con frecuencia operan en silos desconectados de las áreas de sostenibilidad y cumplimiento normativo. Esta fragmentación interna genera puntos ciegos éticos: los sesgos algorítmicos pasan inadvertidos hasta que se materializan en crisis reputacionales o en sanciones legales. La integración efectiva de los principios ESG puede actuar como contrapeso a esta miopía corporativa, en la medida en que subordina los objetivos de optimización técnica a los valores declarados de la organización y a su responsabilidad social ampliada.

La gobernanza de la inteligencia artificial se ha consolidado, en los últimos años, como un campo de estudio dentro de la administración estratégica y la teoría de la organización. A diferencia de las tecnologías informáticas convencionales, los sistemas de aprendizaje automático presentan propiedades de opacidad, autonomía decisoria y escalabilidad que desafían los mecanismos tradicionales de control directivo (Floridi et al., 2018). La literatura sobre gobernanza corporativa, que se remonta a los trabajos seminales de Jensen y Meckling (1976) sobre los costos de agencia, ha evolucionado de un enfoque centrado casi exclusivamente en el control financiero hacia modelos de gestión de riesgos más amplios, que incorporan la supervisión de activos intangibles y de tecnologías complejas cuyo funcionamiento interno resulta difícil de auditar incluso para los propios directores.

2.1 La economía de la predicción y la racionalidad algorítmica

Desde la perspectiva de la economía de la inteligencia artificial, esta tecnología reduce de forma drástica el costo de la predicción (Agrawal et al., 2018). Esta reducción transforma los procesos de toma de decisiones

corporativas, pues permite automatizar decisiones operativas y estratégicas a una velocidad antes impensable. Sin embargo, sustituir la deliberación humana por predicciones algorítmicas introduce sesgos cognitivos y sistémicos difíciles de auditar. La racionalidad algorítmica opera sobre correlaciones estadísticas derivadas de datos históricos, que con frecuencia replican y amplifican patrones de exclusión social y discriminación que ya existían antes de la implementación del sistema. O'Neil (2016) documentó de manera extensa cómo modelos aparentemente técnicos y neutros, aplicados a la evaluación crediticia, la contratación de personal o la vigilancia policial, terminan reproduciendo desigualdades estructurales bajo una apariencia de objetividad matemática que dificulta su cuestionamiento.

Este fenómeno adquiere una relevancia particular en América Latina, donde los conjuntos de datos disponibles para entrenar modelos suelen reflejar desigualdades históricas de acceso al crédito, a la educación formal y al empleo estable. Un sistema de puntuación crediticia entrenado con datos de una población que, durante décadas, tuvo acceso restringido al sistema financiero formal tenderá a perpetuar esa exclusión, incluso si sus desarrolladores no tuvieron intención discriminatoria alguna. Esta observación se conecta con el concepto de "sesgo de disponibilidad de datos" (data availability bias) discutido por Barocas y Selbst (2016), quienes advierten que la calidad ética de un sistema algorítmico depende, en buena medida, de la representatividad histórica de los datos con los que fue entrenado.

2.2 El marco del triple resultado y la integración ESG

La integración de los criterios ESG en la estrategia corporativa responde a la necesidad de gestionar externalidades socioambientales y de garantizar la viabilidad financiera de la empresa en el largo plazo (Eccles, Ioannou y Serafeim, 2014). Siguiendo el enfoque del triple resultado, formulado originalmente por Elkington (1997), la creación de valor sostenible exige que las organizaciones equilibren el desempeño económico con la responsabilidad ambiental y social. La intersección entre la inteligencia artificial y el desempeño ESG constituye un ámbito de estudio relativamente reciente, en el que convergen la ética aplicada, la economía institucional y la teoría de las partes interesadas formulada por Freeman (1984).

Vale la pena detenerse en una precisión conceptual. Los criterios ESG no constituyen un marco monolítico ni universalmente aceptado; conviven distintos estándares de reporte, entre los que destacan los del Global Reporting Initiative (GRI), los de la Sustainability Accounting Standards

Board (SASB) y, más recientemente, los estándares del International Sustainability Standards Board (ISSB), creado en 2021 bajo el paraguas de la Fundación IFRS. Cada uno de estos marcos prioriza de manera distinta la materialidad de los indicadores, lo que introduce cierta ambigüedad al comparar el desempeño ESG entre empresas y sectores. Aplicado a la inteligencia artificial, este problema se agrava, porque ninguno de los estándares mencionados incorpora todavía, de forma sistemática, indicadores específicos sobre gobernanza algorítmica, sesgo de datos o auditoría de modelos, aunque organismos como la Organización para la Cooperación y el Desarrollo Económicos ya han comenzado a avanzar en esa dirección (OCDE, 2019, 2024).

La Recomendación del Consejo sobre Inteligencia Artificial de la OCDE, adoptada en 2019 y actualizada en 2024, propone cinco principios orientadores para el desarrollo de una inteligencia artificial confiable: crecimiento inclusivo y bienestar sostenible; valores centrados en las personas y equidad; transparencia y explicabilidad; robustez, seguridad y protección; y rendición de cuentas. Estos principios, aunque no vinculantes, han influido en marcos regulatorios posteriores, incluida la Ley de Inteligencia Artificial de la Unión Europea, y ofrecen un punto de referencia útil para las corporaciones latinoamericanas que buscan anticiparse a una regulación todavía en construcción en la región.

2.3 Teoría de la agencia y explicabilidad algorítmica

La teoría de la agencia ampliada (Eisenhardt, 1989) sugiere que los directores deben rendir cuentas no solo a los accionistas tradicionales, sino a un conjunto más amplio de partes interesadas afectadas por las decisiones automatizadas. La opacidad de los algoritmos aumenta los costos de transacción y la asimetría informacional entre la empresa y sus públicos, lo que hace necesaria la adopción de estándares rigurosos de explicabilidad y de auditoría externa (Kuziemski y Misuraca, 2020). Esta idea se conecta con el concepto de "caja negra algorítmica" (algorithmic black box), que Burrell (2016) analiza a partir de tres fuentes distintas de opacidad: la opacidad intencional, cuando la empresa oculta deliberadamente el funcionamiento del sistema por razones de propiedad intelectual; la opacidad técnica, derivada de la complejidad matemática de ciertos modelos; y la opacidad de analfabetismo, que surge cuando quienes deben supervisar el sistema carecen de las competencias técnicas necesarias para comprenderlo.

Esta última fuente de opacidad resulta especialmente relevante para los consejos de administración latinoamericanos, en los que la formación

técnica en ciencia de datos suele ser limitada entre los miembros no ejecutivos. Wilson y Daugherty (2018) sostienen que las organizaciones que logran los mejores resultados en la colaboración entre humanos e inteligencia artificial son aquellas que invierten en la capacitación de sus cuadros directivos, de modo que estos puedan formular preguntas pertinentes a los equipos técnicos y ejercer una supervisión sustantiva, y no meramente formal.

2.4 Gobernanza corporativa en América Latina: rasgos distintivos

Para comprender de qué manera la gobernanza de la inteligencia artificial se articula con los principios ESG en la región, es necesario partir de las particularidades de la gobernanza corporativa latinoamericana. La literatura comparada sobre sistemas de gobernanza corporativa ha documentado que, a diferencia de los mercados anglosajones, caracterizados por una propiedad accionaria dispersa, las empresas latinoamericanas presentan estructuras de propiedad altamente concentradas, con frecuencia de carácter familiar, y una separación limitada entre propiedad y control (La Porta et al., 2000). Esta concentración accionaria tiene efectos ambivalentes sobre la gobernanza tecnológica. Por un lado, puede facilitar decisiones más ágiles y una visión de largo plazo menos condicionada por la presión trimestral de los mercados de capitales. Por otro lado, reduce los contrapesos internos y puede favorecer que las decisiones sobre adopción de inteligencia artificial queden concentradas en un número reducido de directores, sin la revisión independiente que ofrecería un consejo con mayor diversidad de intereses.

Husted y Allen (2006) y, más recientemente, Marano y Kostova (2016), estudiaron cómo las prácticas de responsabilidad social corporativa en América Latina se desarrollaron en gran medida como respuesta a presiones institucionales externas, entre ellas la exigencia de inversionistas internacionales y de organismos multilaterales, más que como resultado de una convicción interna consolidada. Esta observación es relevante para el tema que nos ocupa, pues sugiere que la adopción de marcos de gobernanza de inteligencia artificial en la región podría seguir un patrón similar: una adopción inicial de carácter simbólico, orientada a satisfacer requisitos de reporte ESG exigidos por fondos de inversión internacionales, que solo con el tiempo se traduce en cambios sustantivos en las prácticas organizacionales. Esta distinción entre adopción simbólica y adopción sustantiva de prácticas de gobernanza corporativa fue documentada originalmente por Westphal y Zajac (1998) en el contexto de los planes de remuneración ejecutiva, y resulta plenamente aplicable al análisis de los

comités de ética de inteligencia artificial que varias corporaciones latinoamericanas han comenzado a instalar.

2.5 Ética de la inteligencia artificial: del principio a la práctica

La proliferación de códigos de ética de inteligencia artificial en el sector corporativo ha sido notable en los últimos años. Jobin, Ienca y Vayena (2019) analizaron 84 documentos de directrices éticas publicados por gobiernos, empresas y organizaciones internacionales, y encontraron una convergencia relativamente alta en torno a cinco principios: transparencia, justicia y equidad, no maleficencia, responsabilidad y privacidad. Sin embargo, los mismos autores advierten que esa convergencia discursiva contrasta con una divergencia sustancial en la manera en que cada principio se interpreta y se implementa en la práctica, lo que abre espacio para lo que la literatura ha denominado "ethics washing", es decir, el uso de un lenguaje ético como cobertura reputacional sin cambios operativos correspondientes (Wagner, 2018).

Rahwan (2018) propone el concepto de "sociedad en el circuito" (society-in-the-loop) para describir un modelo de gobernanza en el que los sistemas algorítmicos no son auditados únicamente por expertos técnicos, sino que incorporan mecanismos de participación de las comunidades afectadas por sus decisiones. Esta perspectiva resulta particularmente pertinente para el contexto latinoamericano, donde la desconfianza histórica hacia las instituciones económicas hace que la legitimidad de un sistema de gobernanza dependa, en buena medida, de que las partes interesadas perciban que tienen algún grado de voz real en su diseño y en su supervisión, y no solo en la comunicación posterior de sus resultados.

2.6 El panorama regulatorio emergente en América Latina

Aunque la región carece de una legislación integral sobre inteligencia artificial comparable a la Ley de Inteligencia Artificial de la Unión Europea, en los últimos años varios países han comenzado a avanzar en esa dirección, con resultados desiguales. Brasil discute desde 2021 el Proyecto de Ley 2338, que propone un marco de clasificación de sistemas de inteligencia artificial según su nivel de riesgo, inspirado explícitamente en el modelo europeo, y que incluye disposiciones específicas sobre auditoría algorítmica y sobre responsabilidad civil de los proveedores de tecnología. Chile aprobó en 2024 una política nacional de inteligencia artificial que, aunque no tiene carácter vinculante, establece principios orientadores similares a los de la OCDE. Perú promulgó en 2023 una ley que declara de interés nacional el desarrollo de la inteligencia artificial, aunque su reglamentación sigue

pendiente en varios aspectos centrales. México, por su parte, todavía no cuenta con una ley federal específica, y el debate regulatorio se ha concentrado principalmente en el ámbito de la protección de datos personales.

Esta heterogeneidad regulatoria genera un problema de coordinación para las corporaciones multilatinas, es decir, aquellas empresas de origen latinoamericano que operan en varios países de la región simultáneamente. Una misma práctica de uso de inteligencia artificial puede estar sujeta a obligaciones de transparencia distintas según el país en que se aplique, lo que complica el diseño de políticas corporativas unificadas. Ante esta fragmentación, algunas empresas han optado por adoptar como estándar interno el marco regulatorio más exigente entre los países en que operan, una estrategia conocida en la literatura de cumplimiento normativo como "nivelación por arriba" (leveling up), que reduce el riesgo de incumplimiento en cualquiera de las jurisdicciones y, de paso, simplifica la gestión interna de la política de gobernanza tecnológica.

Esta laguna regulatoria parcial refuerza el argumento central de este trabajo: en ausencia de una obligación legal homogénea y exigible en toda la región, la responsabilidad de establecer estándares rigurosos de gobernanza algorítmica recae, en gran medida, sobre los propios consejos de administración y sobre la disciplina que los mercados de capitales internacionales puedan imponer mediante sus criterios de inversión responsable.

En síntesis, el marco teórico que sustenta este trabajo integra cinco cuerpos de literatura: la economía de la predicción y la racionalidad algorítmica, el marco del triple resultado y los estándares ESG, la teoría de la agencia aplicada a la explicabilidad algorítmica, la literatura sobre gobernanza corporativa y ética de la inteligencia artificial en economías emergentes, y el panorama regulatorio emergente propio de la región. A partir de esta integración, se construye, en las secciones siguientes, un modelo que articula estas dimensiones para el análisis del caso latinoamericano.

3 Metodología

Este estudio adopta una metodología de revisión teórica de carácter integrativo (Torraco, 2016), orientada a sintetizar literatura científica interdisciplinaria sobre gobernanza algorítmica, ética corporativa y desempeño ESG. Se optó por este diseño, y no por una revisión sistemática en sentido estricto, porque el objetivo del trabajo no es cuantificar la frecuencia de determinados hallazgos empíricos, sino construir un marco

conceptual capaz de articular cuerpos de literatura que, hasta ahora, se han desarrollado de manera relativamente independiente: la gobernanza corporativa, la ética de la inteligencia artificial y los estudios de sostenibilidad empresarial en economías emergentes.

El proceso metodológico se estructuró en cuatro fases, diseñadas para garantizar la trazabilidad analítica del trabajo.

En la primera fase, se realizó un relevamiento bibliográfico centrado en marcos regulatorios, directrices éticas y modelos de gobernanza de inteligencia artificial, tanto en economías emergentes como en economías desarrolladas, utilizando bases de datos académicas internacionales, entre ellas Scopus, Web of Science y Google Scholar. Se priorizaron artículos publicados en revistas indexadas de administración, ética aplicada y economía institucional, entre ellas *Journal of Business Ethics*, *Academy of Management Review*, *Strategic Management Journal*, *Corporate Governance: An International Review*, *Business Ethics Quarterly* y *Minds and Machines*, además de documentos técnicos de organismos multilaterales como la OCDE y la UNESCO.

En la segunda fase, se contrastaron los objetivos declarados de la transformación digital corporativa con los requisitos normativos de los principales estándares internacionales de reporte ESG, entre ellos el GRI, el SASB y el ISSB. Este contraste permitió identificar las áreas en que los marcos de sostenibilidad existentes ya contemplan, aunque de manera incipiente, criterios aplicables a la gobernanza tecnológica, y aquellas en que persiste una laguna normativa evidente.

En la tercera fase, se identificaron los puntos de fricción, de sinergia y las paradojas estratégicas entre ambos dominios, mediante un ejercicio de codificación temática de la literatura revisada. Este ejercicio se apoyó en la técnica de análisis de contenido cualitativo descrita por Elo y Kyngäs (2008), que permite organizar categorías conceptuales emergentes a partir de la lectura sistemática de un corpus documental.

Por último, en la cuarta fase, se sintetizaron los hallazgos para estructurar un modelo integrador de gobernanza corporativa aplicable al contexto latinoamericano, presentado y discutido en las secciones cuarta, quinta y sexta de este artículo. Como toda revisión integrativa, este trabajo no pretende agotar la literatura disponible sobre el tema, sino ofrecer una síntesis argumentada capaz de orientar tanto la reflexión académica como la práctica directiva.

4 Intersecciones críticas entre la inteligencia artificial y los pilares ESG

El análisis integrativo realizado en este estudio muestra que la inteligencia artificial impacta cada uno de los tres pilares ESG de manera diferenciada, lo que exige respuestas de gobernanza específicas y, al mismo tiempo, coordinadas entre sí. A continuación se examina cada dimensión por separado, sin perder de vista que, en la práctica organizacional, estas tres dimensiones interactúan de forma constante.

4.1 Dimensión ambiental (E) y sostenibilidad operativa

La inteligencia artificial ofrece herramientas de considerable potencial para la modelización climática, la optimización de redes eléctricas, la gestión eficiente de recursos hídricos y la reducción de emisiones industriales. Empresas de sectores intensivos en energía han comenzado a usar modelos predictivos para optimizar el consumo energético de sus plantas y reducir su huella de carbono. Sin embargo, esta misma tecnología tiene una huella ecológica operativa considerable, asociada al consumo intensivo de energía y de agua en los centros de datos de gran escala necesarios para entrenar y operar los modelos de aprendizaje profundo. Strubell, Ganesh y McCallum (2019) documentaron que el entrenamiento de un único modelo de procesamiento de lenguaje natural de gran tamaño puede generar emisiones de dióxido de carbono comparables a las de varios automóviles a lo largo de toda su vida útil, dato que ha sido citado con frecuencia en la literatura posterior sobre sostenibilidad digital.

Esta contradicción ambiental, entre el potencial de la inteligencia artificial para mitigar el cambio climático y su propio costo energético, obliga a las corporaciones a auditar de manera activa todo el ciclo de vida de sus infraestructuras de cómputo, y no solo el efecto ambiental de las aplicaciones finales de la tecnología. En América Latina, donde varios países todavía dependen en gran medida de fuentes de energía no renovable para su matriz eléctrica, la instalación de nuevos centros de datos presenta un dilema adicional: la ubicación geográfica de estas infraestructuras puede determinar si su huella de carbono es o no compensada por energías limpias disponibles localmente. Algunas corporaciones de la región han comenzado a exigir a sus proveedores de servicios en la nube información desagregada sobre el origen de la energía utilizada, práctica que aún no es generalizada, pero que apunta en la dirección correcta.

4.2 Dimensión social (S) y equidad algorítmica

Los sistemas automatizados de toma de decisiones aplicados a la gestión de recursos humanos, a la evaluación de riesgo crediticio o a la atención al cliente pueden incorporar sesgos históricos discriminatorios presentes en los datos con los que fueron entrenados, lo que afecta negativamente la equidad social y la inclusión de poblaciones vulnerables en América Latina. Este riesgo no es meramente hipotético. Estudios realizados en Estados Unidos sobre sistemas de puntuación crediticia y de reconocimiento facial demostraron tasas de error significativamente más altas para personas de piel oscura y para mujeres, en comparación con hombres de piel clara (Buolamwini y Gebru, 2018). Aunque la evidencia empírica específica para la región latinoamericana todavía es escasa, hay razones para pensar que fenómenos similares podrían ocurrir, dada la superposición histórica entre desigualdad racial, desigualdad de género y desigualdad de acceso a servicios financieros formales en varios países de la región.

Un ámbito particularmente sensible es el de la selección de personal mediante sistemas automatizados. Empresas de distintos sectores han comenzado a usar algoritmos para filtrar currículums, analizar entrevistas en video o evaluar el desempeño de sus empleados. Cuando estos sistemas se entrenan con datos históricos de contratación que reflejan patrones anteriores de discriminación, ya sea por edad, por género o por origen socioeconómico, corren el riesgo de perpetuar esos mismos patrones bajo una apariencia de neutralidad técnica. Raji y Buolamwini (2019) documentaron que la publicación de auditorías externas sobre el desempeño diferencial de estos sistemas, desagregado por grupos demográficos, puede incentivar a las empresas desarrolladoras a corregir sus modelos, lo que sugiere que la transparencia, más que la simple prohibición regulatoria, puede ser una palanca de cambio efectiva en contextos donde la capacidad de fiscalización estatal es limitada, como ocurre en buena parte de América Latina.

4.3 Dimensión de gobernanza (G) y transparencia institucional

La opacidad de los algoritmos de aprendizaje automático debilita los mecanismos tradicionales de rendición de cuentas ante accionistas, reguladores y sociedad civil, lo que exige nuevas formas de auditoría algorítmica independiente y de explicabilidad técnica (Raji et al., 2020). Esta dimensión se conecta directamente con la literatura clásica de gobernanza corporativa, en la medida en que la introducción de sistemas de decisión automatizada añade una capa adicional de complejidad a los ya conocidos problemas de agencia entre directores, accionistas y otras partes interesadas.

Un problema particular de la dimensión de gobernanza en la región es la escasa presencia de perfiles técnicos especializados en tecnología dentro de los consejos de administración. A diferencia de lo que ocurre en mercados más desarrollados, donde algunas empresas ya cuentan con comités específicos de tecnología o con consejeros independientes con formación en ciencia de datos, en América Latina esta práctica todavía es poco frecuente. Esta carencia limita la capacidad efectiva de supervisión del consejo sobre las decisiones relacionadas con la inteligencia artificial, y traslada, en la práctica, buena parte de la responsabilidad de gobernanza a los propios equipos técnicos que desarrollan o adquieren los sistemas, lo que reproduce precisamente el problema de la opacidad de analfabetismo descrito por Burrell (2016) y mencionado en la sección anterior.

4.4 Interacciones entre pilares: sinergias y tensiones

Conviene destacar que las tres dimensiones ESG no operan de manera aislada frente a la inteligencia artificial, sino que interactúan entre sí, unas veces de forma sinérgica, otras de forma conflictiva. Un ejemplo de sinergia es el uso de modelos predictivos para optimizar rutas logísticas, lo que puede reducir simultáneamente el consumo de combustible, mejorar las condiciones de trabajo de los conductores al disminuir jornadas excesivas, y facilitar el reporte transparente de indicadores de eficiencia a los inversionistas. Un ejemplo de tensión, a su vez, es el uso de sistemas de vigilancia algorítmica del desempeño laboral que, aunque puedan mejorar indicadores de productividad reportables bajo criterios de gobernanza, generan al mismo tiempo efectos negativos sobre el bienestar y la privacidad de los trabajadores, es decir, sobre la dimensión social.

Esta observación es coherente con la literatura sobre disyuntivas (trade-offs) entre dimensiones ESG, que señala que la maximización simultánea de las tres dimensiones no siempre es posible, y que las organizaciones frecuentemente necesitan tomar decisiones que privilegian una dimensión sobre otra según su contexto específico (Hahn, Figge, Pinkse y Preuss, 2018). Aplicado a la gobernanza de la inteligencia artificial, este marco sugiere que las corporaciones latinoamericanas necesitan mecanismos explícitos de deliberación interna que les permitan identificar y resolver, de manera consciente y documentada, estas tensiones, en lugar de dejar que se resuelvan de hecho según la lógica de eficiencia que suele predominar entre los equipos técnicos.

5 Mecanismos institucionales para la sostenibilidad algorítmica

Para mitigar los riesgos descritos en la sección anterior, las corporaciones necesitan institucionalizar mecanismos internos de control que aseguren la alineación entre las implementaciones tecnológicas y sus compromisos ESG. A partir de la revisión de literatura realizada, se identifican cinco mecanismos que, en conjunto, constituyen la base de un modelo de gobernanza algorítmica orientado a la sostenibilidad.

5.1 Comités de ética de inteligencia artificial con poder de veto

El primer mecanismo es la creación de comités de ética de inteligencia artificial dotados de autoridad real, incluido el poder de veto sobre proyectos automatizados de alto riesgo. La literatura sobre comités de ética corporativa advierte, sin embargo, que su eficacia depende críticamente de su composición y de su posición jerárquica dentro de la organización. Un comité compuesto exclusivamente por personal del área de tecnología, o que reporta directamente al director de tecnología, difícilmente ejercerá una supervisión independiente. Por el contrario, los comités más eficaces suelen incluir representantes de las áreas de cumplimiento normativo, recursos humanos y sostenibilidad y, cuando es posible, miembros externos con formación en ética aplicada o en derechos digitales, y reportan directamente al consejo de administración o a un comité del consejo (Rességuier y Rodrigues, 2020).

5.2 Auditorías periódicas de equidad y sesgo

El segundo mecanismo consiste en la realización periódica de auditorías de equidad y de sesgo sobre los sistemas algorítmicos en producción, y no solo en la etapa de diseño. Raji et al. (2020) proponen un marco de auditoría interna de extremo a extremo que documenta cada etapa del ciclo de vida de un modelo, desde la recolección de datos hasta el monitoreo posterior a su implementación, con el objetivo de identificar puntos de fricción ética antes de que se traduzcan en daños concretos. Este tipo de auditoría exige, sin embargo, capacidades técnicas especializadas que muchas corporaciones latinoamericanas de tamaño medio todavía no poseen internamente, lo que abre una oportunidad de mercado para empresas de auditoría algorítmica independiente, un sector todavía incipiente en la región, pero en expansión.

5.3 Informes de transparencia algorítmica

El tercer mecanismo es la publicación de informes de transparencia algorítmica dirigidos a las partes interesadas, que documenten de manera accesible, y no meramente técnica, cómo la organización utiliza sistemas de inteligencia artificial en sus decisiones más sensibles. La literatura sobre

comunicación corporativa de sostenibilidad advierte que estos informes corren el riesgo de convertirse en ejercicios de relaciones públicas si no van acompañados de indicadores verificables externamente (Cho, Laine, Roberts y Rodrigue, 2015). Por esta razón, se recomienda que los informes de transparencia algorítmica se integren a los reportes ESG ya existentes, y que se sometan a los mismos procesos de verificación externa aplicados, por ejemplo, a los indicadores de emisiones de carbono.

5.4 Capacitación y alfabetización algorítmica del consejo de administración

El cuarto mecanismo, con frecuencia subestimado en la literatura, es la capacitación específica de los miembros del consejo de administración y de la alta dirección en conceptos básicos de inteligencia artificial y de ética de datos. Como se discutió en la sección teórica, la opacidad de analfabetismo constituye una barrera significativa para la supervisión efectiva. Los programas de capacitación continua, complementados por la incorporación gradual de consejeros independientes con formación técnica, pueden reducir esta brecha y fortalecer la capacidad del consejo para formular preguntas pertinentes y exigir explicaciones comprensibles a los equipos técnicos.

5.5 Mecanismos de participación de partes interesadas externas

El quinto mecanismo, inspirado en la propuesta de sociedad en el circuito de Rahwan (2018), consiste en incorporar canales formales de consulta con las comunidades y los grupos afectados por los sistemas algorítmicos de la empresa, particularmente en sectores como servicios financieros, salud o empleo. Estos canales pueden adoptar la forma de paneles consultivos externos, encuestas periódicas o mesas de diálogo con organizaciones de la sociedad civil especializadas en derechos digitales. Aunque esta práctica todavía es poco común entre las corporaciones latinoamericanas, algunas experiencias en el sector financiero de la región muestran que incorporar este tipo de mecanismo puede anticipar conflictos regulatorios y mejorar la legitimidad social de los proyectos de automatización antes de su implementación completa.

La combinación de estos cinco mecanismos, comité de ética con poder de veto, auditorías periódicas de equidad, informes de transparencia verificables, capacitación del consejo y participación de partes interesadas externas, constituye lo que este trabajo denomina un modelo de gobernanza algorítmica orientado a la sostenibilidad. Se trata de un modelo exigente, que requiere inversión de tiempo y de recursos, pero que es coherente con la evidencia disponible sobre los factores que distinguen la adopción

simbólica de la adopción sustantiva de prácticas de gobernanza corporativa (Westphal y Zajac, 1998).

4.5 El caso del sector financiero como referencia regional

El sector financiero latinoamericano ofrece un buen punto de referencia para observar, de manera concreta, cómo se manifiestan estas tensiones en la práctica. Los bancos y las plataformas de crédito digital de la región adoptaron rápidamente modelos de puntuación crediticia basados en aprendizaje automático, que incorporan variables alternativas a las tradicionales, como el comportamiento en redes sociales, los patrones de uso del teléfono móvil o el historial de pago de servicios básicos. Esta innovación permitió, en varios casos documentados, ampliar el acceso al crédito formal para poblaciones históricamente excluidas del sistema financiero, lo que representa una contribución positiva a la dimensión social de la sostenibilidad corporativa.

Sin embargo, la misma innovación introduce riesgos que no siempre son evidentes a primera vista. El uso de variables alternativas, como el consumo de datos móviles o la geolocalización, puede terminar funcionando como un indicador indirecto (proxy) de variables socioeconómicas que la regulación de protección de datos prohíbe utilizar directamente, como el nivel de ingresos o el lugar de residencia. Este fenómeno, conocido en la literatura como discriminación por proxy, ha sido documentado en mercados desarrollados, y existen razones sólidas para pensar que también ocurre en América Latina, aunque la evidencia empírica específica todavía es escasa. La regulación financiera de la región, históricamente centrada en el riesgo de solvencia y en el lavado de activos, no siempre cuenta todavía con los instrumentos técnicos necesarios para fiscalizar este tipo de riesgo algorítmico, lo que refuerza, una vez más, la importancia de la autorregulación corporativa discutida en secciones anteriores.

Superintendencias financieras de varios países de la región, entre ellas las de Colombia y México, han comenzado a emitir orientaciones sobre el uso responsable de la inteligencia artificial en la evaluación de riesgo crediticio, aunque su carácter sigue siendo, en su mayoría, indicativo y no vinculante. Este ejemplo sectorial ilustra con claridad el argumento general de este trabajo: la tecnología puede contribuir de forma genuina a los objetivos ESG de una organización, pero solo si su implementación va acompañada de mecanismos de auditoría y de rendición de cuentas capaces de detectar y corregir los efectos indeseados que, de otro modo, permanecerían invisibles hasta convertirse en un problema reputacional o regulatorio de gran escala.

6 Implicaciones estratégicas para la alta dirección

Los resultados de este ensayo sugieren que la alta dirección corporativa en América Latina necesita superar la visión puramente instrumental de la tecnología, según la cual la inteligencia artificial se evalúa exclusivamente en función de sus ganancias de eficiencia y de su retorno financiero inmediato. La gobernanza de la inteligencia artificial, integrada de manera efectiva con los principios ESG, representa un factor de diferenciación competitiva y de legitimidad institucional en mercados globales cada vez más regulados y más atentos a los riesgos reputacionales asociados a la tecnología.

En primer lugar, la evidencia revisada sugiere que las empresas que se anticipan a los estándares regulatorios internacionales, en lugar de esperar a que la legislación local los vuelva obligatorios, obtienen ventajas reputacionales frente a inversionistas institucionales que ya incorporan criterios de gobernanza tecnológica en sus procesos de debida diligencia. Los grandes fondos de inversión han comenzado a incluir preguntas específicas sobre gobernanza de inteligencia artificial en sus cuestionarios de evaluación ESG, lo que convierte este tema, para las empresas que buscan financiamiento internacional, en una cuestión de acceso al capital, y no solo de ética corporativa.

En segundo lugar, integrar la gobernanza de la inteligencia artificial con los principios ESG exige redefinir las responsabilidades dentro de la estructura directiva. En muchas organizaciones latinoamericanas, la responsabilidad sobre la inteligencia artificial recae exclusivamente en el área de tecnología, mientras que la responsabilidad sobre ESG recae en un área de sostenibilidad separada, con poca interacción entre ambas. Este trabajo sugiere que la creación de instancias de coordinación formal, como el comité de ética de inteligencia artificial descrito en la sección anterior, resulta indispensable para superar esta fragmentación funcional y evitar que los puntos ciegos éticos ya mencionados se traduzcan en crisis reputacionales evitables.

En tercer lugar, la alta dirección debe considerar que la gestión de riesgos algorítmicos no puede improvisarse una vez que un sistema ya está en producción. La literatura sobre gestión de riesgos de tecnologías emergentes recomienda incorporar evaluaciones de impacto ético desde las etapas iniciales del diseño de cualquier proyecto de inteligencia artificial, de manera análoga a cómo se realizan las evaluaciones de impacto ambiental en proyectos de infraestructura física (Reisman, Schultz, Crawford y

Whittaker, 2018). Esta práctica, todavía poco difundida en la región, permitiría identificar riesgos de sesgo, de privacidad y de exclusión social antes de que un sistema alcance su fase de implementación masiva, momento en el que corregirlo se vuelve considerablemente más costoso.

Por último, conviene señalar que la inversión en gobernanza algorítmica no debe entenderse como un costo puramente defensivo. Las empresas que han desarrollado capacidades sólidas de auditoría y de explicabilidad de sus sistemas suelen estar mejor posicionadas para escalar nuevas aplicaciones de inteligencia artificial con mayor rapidez, porque cuentan con protocolos ya probados para evaluar riesgos y obtener la aprobación interna necesaria. En este sentido, la gobernanza rigurosa puede funcionar, más que como un freno, como una capacidad organizacional que permite una innovación más rápida y, al mismo tiempo, más responsable.

Un elemento adicional que merece la atención de la alta dirección es la relación entre la gobernanza de la inteligencia artificial y la retención de talento especializado. Los profesionales jóvenes de ciencia de datos e ingeniería de software, en particular quienes se formaron en universidades con programas de ética tecnológica, muestran una sensibilidad creciente hacia el uso responsable de la tecnología que desarrollan. Encuestas realizadas entre profesionales del sector en distintos países han mostrado que una proporción significativa de estos trabajadores considera la ausencia de políticas claras de ética de datos un factor relevante al evaluar una oferta de empleo o al decidir si permanecer en una organización. Para las corporaciones latinoamericanas, que ya enfrentan una competencia intensa por talento técnico calificado frente a empresas globales que ofrecen condiciones de trabajo remoto y salarios en moneda extranjera, contar con una política de gobernanza algorítmica creíble puede convertirse en un argumento adicional de atracción y retención de personal, un beneficio raramente contabilizado en los análisis de retorno sobre la inversión de estos programas, pero coherente con la evidencia disponible sobre el vínculo entre propósito organizacional y compromiso laboral.

Por último, la alta dirección debe considerar el papel de los órganos de control interno y de auditoría en este proceso. La función de auditoría interna, tradicionalmente centrada en el control financiero y el cumplimiento normativo, necesita ampliar su alcance para incorporar la revisión periódica de los sistemas algorítmicos de mayor riesgo. Esto implica, entre otras cosas, capacitar a los equipos de auditoría en conceptos básicos de aprendizaje automático y establecer protocolos de trabajo conjunto con las áreas de tecnología, de modo que la auditoría no dependa

exclusivamente de la información que los propios equipos técnicos decidan compartir. Sin esta ampliación de alcance, sigue siendo alto el riesgo de que la gobernanza algorítmica quede reducida a una declaración de principios sin verificación efectiva.

7 Consideraciones finales

La gobernanza de la inteligencia artificial y su articulación con los principios ESG configuran, hoy, un tema central para las corporaciones latinoamericanas. Como se ha argumentado a lo largo de este trabajo, la tecnología por sí sola no garantiza sostenibilidad alguna; requiere un marco institucional de gobernanza rigurosa que subordine la eficiencia algorítmica al bienestar social y ambiental de las comunidades en las que opera la empresa. El modelo propuesto, articulado en torno a cinco mecanismos institucionales, ofrece un punto de partida para que los consejos de administración y los equipos directivos puedan avanzar en esa dirección de manera estructurada, y no meramente reactiva frente a presiones regulatorias o a crisis reputacionales puntuales.

A lo largo del artículo se mostró que la inteligencia artificial impacta de manera diferenciada cada uno de los tres pilares ESG. En la dimensión ambiental, el potencial de la tecnología para optimizar procesos convive con una huella energética considerable que las empresas necesitan auditar de forma activa. En la dimensión social, los sistemas automatizados pueden perpetuar desigualdades históricas si no se someten a auditorías de equidad rigurosas y periódicas. En la dimensión de gobernanza, la opacidad de los algoritmos exige nuevas capacidades de supervisión que buena parte de los consejos de administración de la región todavía no ha desarrollado. Ante este panorama, la propuesta central del trabajo es que ninguna de estas tres dimensiones puede gestionarse de forma aislada, y que su integración efectiva depende de mecanismos de coordinación interna explícitos, respaldados por autoridad real dentro de la estructura organizacional.

7.1 Limitaciones del estudio

A pesar de la consistencia conceptual del modelo propuesto, este estudio presenta limitaciones que conviene reconocer con claridad. Por tratarse de una revisión teórico-integrativa, el trabajo no incorpora evidencia empírica cuantitativa ni estudios de caso sectoriales específicos sobre corporaciones latinoamericanas, de modo que el marco propuesto requiere validación en entornos corporativos reales de la región. Además, la literatura revisada proviene, en su mayoría, de contextos institucionales distintos al latinoamericano, principalmente de Estados Unidos y de Europa, lo que

exige cierta cautela al extrapolar hallazgos originados en esos contextos a realidades regulatorias, culturales y económicas diferentes. Por último, dada la velocidad con que evolucionan tanto la tecnología de inteligencia artificial como la regulación asociada a ella, algunos de los marcos normativos discutidos en este trabajo, en particular los relacionados con los estándares de reporte ESG, podrían modificarse de manera significativa en los próximos años.

7.2 Agenda de investigaciones futuras

Como orientación para futuras agendas de investigación, se recomienda el desarrollo de estudios empíricos cuantitativos, mediante modelización de ecuaciones estructurales u otras técnicas equivalentes, que permitan medir el efecto mediador de la gobernanza de inteligencia artificial sobre el desempeño ESG en sectores industriales intensivos en datos, como el financiero, el de telecomunicaciones y el minorista. También sería valioso realizar estudios de caso comparados entre corporaciones latinoamericanas que han implementado comités de ética de inteligencia artificial y aquellas que no lo han hecho, con el objetivo de identificar de manera más precisa los factores organizacionales que facilitan, o que dificultan, la adopción sustantiva, y no meramente simbólica, de estos mecanismos. Por último, dada la escasez de evidencia empírica específica sobre sesgo algorítmico en poblaciones latinoamericanas, se recomienda impulsar investigaciones que documenten de manera sistemática el desempeño diferencial de los sistemas de puntuación crediticia, de selección de personal y de evaluación de riesgo aplicados en la región, siguiendo metodologías similares a las empleadas por Buolamwini y Gebru (2018) en el contexto estadounidense.

7.3 Posicionamiento del autor

Desde la perspectiva del autor, se sostiene que la adopción acrítica de la inteligencia artificial, bajo la retórica genérica de la modernización corporativa, representa una amenaza para la cohesión social en América Latina si no va acompañada de mecanismos de gobernanza sólidos y verificables. La tecnología no es un factor exógeno ni neutro: es el resultado de decisiones institucionales y políticas concretas, que deben gobernarse de manera transparente y, en la medida de lo posible, con la participación efectiva de las comunidades afectadas por sus consecuencias. El desafío para las corporaciones de la región no consiste en elegir entre innovación tecnológica y responsabilidad social, sino en construir las capacidades institucionales necesarias para que ambas avancen en conjunto.

REFERENCIAS

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Press.

Aguerre, C. (2021). Género y gobernanza de la inteligencia artificial en América Latina. *Revista CIDOB d'Afers Internacionals*, (127), 87-108.

Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Addison-Wesley.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671-732. <https://doi.org/10.15779/Z38BG31>

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77-91.

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1-12. <https://doi.org/10.1177/2053951715622512>

Cascio, W. F., & Boudreau, J. W. (2016). The search for global competence: From international HR to strategic agility. *Human Resource Management Review*, 26(1), 5-15. <https://doi.org/10.1016/j.hrmr.2015.09.002>

Cho, C. H., Laine, M., Roberts, R. W., & Rodrigue, M. (2015). Organized hypocrisy, organizational façades, and sustainability reporting. *Accounting, Organizations and Society*, 40, 78-94. <https://doi.org/10.1016/j.aos.2014.12.003>

Cihon, P., Maas, M. M., & Kemp, L. (2020). Should artificial intelligence governance be centralised? Design lessons from history. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 228-234). ACM. <https://doi.org/10.1145/3375627.3375857>

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, 60(11), 2835-2857. <https://doi.org/10.1287/mnsc.2014.1984>

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 57-74. <https://doi.org/10.5465/amr.1989.4279003>

Elkington, J. (1997). *Cannibals with forks: The triple bottom line of 21st century business*. Capstone.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107-115. <https://doi.org/10.1111/j.1365-2648.2007.04569.x>

Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.

Gasser, U., & Almeida, V. A. (2017). Dialogues on artificial intelligence: Governing through code and beyond. *IEEE Internet Computing*, 21(4), 4-11. <https://doi.org/10.1109/MIC.2017.2911432>

Hahn, T., Figge, F., Pinkse, J., & Preuss, L. (2018). A paradox perspective on corporate sustainability: Descriptive, instrumental, and normative aspects. *Journal of Business Ethics*, 148(2), 235-248. <https://doi.org/10.1007/s10551-017-3587-2>

Husted, B. W., & Allen, D. B. (2006). Corporate social responsibility in the multinational enterprise: Strategic and institutional approaches. *Journal of International Business Studies*, 37(6), 838-849. <https://doi.org/10.1057/palgrave.jibs.8400227>

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305-360. [https://doi.org/10.1016/0304-405X\(76\)90026-X](https://doi.org/10.1016/0304-405X(76)90026-X)

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399. <https://doi.org/10.1038/s42256-019-0088-2>

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making. *Telecommunications Policy*, 44(6), 101976. <https://doi.org/10.1016/j.telpol.2020.101976>

La Porta, R., López-de-Silanes, F., Shleifer, A., & Vishny, R. W. (2000). Investor protection and corporate governance. *Journal of Financial Economics*, 58(1-2), 3-27. [https://doi.org/10.1016/S0304-405X\(00\)00065-9](https://doi.org/10.1016/S0304-405X(00)00065-9)

Marano, V., & Kostova, T. (2016). Unpacking the institutional complexity in adoption of CSR practices in multinational enterprises. *Journal of Management Studies*, 53(1), 28-54. <https://doi.org/10.1111/joms.12124>

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-21. <https://doi.org/10.1177/2053951716679679>

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

Organisation for Economic Co-operation and Development. (2019). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.

Organisation for Economic Co-operation and Development. (2024). Recommendation of the Council on Artificial Intelligence (amended). OECD/LEGAL/0449.

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, 20(1), 5-14. <https://doi.org/10.1007/s10676-017-9430-8>

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 429-435). ACM. <https://doi.org/10.1145/3306618.3314244>

- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency (pp. 33-44). ACM.
<https://doi.org/10.1145/3351095.3372842>
- Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic impact assessments: A practical framework for public agency accountability. AI Now Institute.
- Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*, 7(2), 1-5. <https://doi.org/10.1177/2053951720942541>
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645-3650). ACL. <https://doi.org/10.18653/v1/P19-1355>
- Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13), 1319-1350.
<https://doi.org/10.1002/smj.640>
- Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, 15(4), 404-428.
<https://doi.org/10.1177/1534484316671606>
- Wagner, B. (2018). Ethics as an escape from regulation: From ethics-washing to ethics-shopping? In E. Bayamlioğlu, I. Baraliuc, L. Janssens, & M. Hildebrandt (Eds.), *Being profiled: Cogitas ergo sum* (pp. 84-89). Amsterdam University Press.
- Westphal, J. D., & Zajac, E. J. (1998). The symbolic management of stockholders: Corporate governance reforms and shareholder reactions. *Administrative Science Quarterly*, 43(1), 127-153.
<https://doi.org/10.2307/2393593>
- Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 195-200). ACM. <https://doi.org/10.1145/3306618.3314289>
- Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, 96(4), 114-123.
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector: Applications and challenges. *International Journal of Public Administration*, 42(7), 596-615.
<https://doi.org/10.1080/01900692.2018.1498103>
- Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.

CONTRIBUCIÓN DEL AUTOR

Conceptualización, metodología, investigación, análisis formal, curaduría de datos, redacción del manuscrito original y revisión y edición.

NOTA BIOGRÁFICA

Lincon Lopes está vinculado al Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), São Paulo, SP, Brasil.

APROBACIÓN ÉTICA

No aplica. El estudio es una revisión teórica integrativa de literatura publicada, que no involucra la recolección de datos de seres humanos ni de animales.

DISPONIBILIDAD DE LOS DATOS

No aplica. El artículo no produjo un conjunto de datos primarios.

DECLARACIONES EDITORIALES

Financiamiento: el estudio no recibió financiamiento externo.

Conflicto de intereses: el autor declara no tener conflicto de intereses.

Uso de inteligencia artificial: el autor no declaró el uso de herramientas de inteligencia artificial.