

Modelo de Empresa Inteligente y Sostenible: Integración de IA, CSR y ESG

Smart and Sustainable Enterprise: Strategic Integration of AI, CSR, and ESG in Emerging Economies

Dr. José Fernando Hernández González; Dra. María Esther Fragoso Terán y
Dr. Arturo López Saldiña

Resumen. La incorporación acelerada de la inteligencia artificial en las organizaciones contemporáneas ha generado nuevas tensiones entre innovación tecnológica, responsabilidad social corporativa y legitimidad institucional. Este artículo propone un modelo de empresa inteligente y sostenible mediante la integración ética de la inteligencia artificial, la responsabilidad social corporativa y los criterios ambientales, sociales y de gobernanza. Desde un enfoque conceptual, crítico y propositivo, se analizan las limitaciones del *business case* como fundamento dominante de la responsabilidad corporativa, especialmente frente a riesgos como opacidad algorítmica, sesgo, vigilancia, desplazamiento laboral, oportunismo ESG y *machine washing*. El estudio sostiene que la sostenibilidad empresarial en la era digital no puede reducirse a la automatización de indicadores ni a declaraciones reputacionales sobre IA responsable. Por el contrario, requiere una gobernanza ética capaz de traducir principios normativos en mecanismos verificables de auditoría, explicabilidad, trazabilidad, supervisión humana y rendición de cuentas. El argumento humano se plantea como eje normativo para subordinar la eficiencia algorítmica a la dignidad humana, la justicia social y la legitimidad organizacional. Se concluye que una empresa verdaderamente inteligente y sostenible no es aquella que adopta más tecnología, sino aquella que demuestra que sus sistemas inteligentes operan bajo responsabilidad social, transparencia institucional y sostenibilidad integral.

Palabras clave: inteligencia artificial, responsabilidad social corporativa, ESG, empresa inteligente, gobernanza ética, *machine washing*, sostenibilidad corporativa.

Abstract. The accelerated incorporation of artificial intelligence into contemporary organizations has generated new tensions among technological innovation, corporate social responsibility, and institutional legitimacy. This article proposes a smart and sustainable enterprise model through the ethical integration of artificial intelligence, corporate social responsibility, and environmental, social, and governance criteria. From a conceptual, critical, and propositional approach, the study analyzes the limitations of the *business case* as the dominant foundation of corporate responsibility, particularly in relation to risks such as algorithmic opacity, bias, surveillance, labor displacement, ESG

opportunism, and machine washing. The article argues that business sustainability in the digital era cannot be reduced to the automation of indicators or reputational statements about responsible AI. Instead, it requires ethical governance capable of translating normative principles into verifiable mechanisms of auditing, explainability, traceability, human oversight, and accountability. The human argument is presented as a normative axis for subordinating algorithmic efficiency to human dignity, social justice, and organizational legitimacy. The article concludes that a truly smart and sustainable enterprise is not the one that adopts more technology, but the one that demonstrates that its intelligent systems operate under social responsibility, institutional transparency, and integral sustainability.

Keywords: artificial intelligence, corporate social responsibility, ESG, smart enterprise, ethical governance, machine washing, corporate sustainability.

Introducción

La expansión acelerada de la inteligencia artificial en las organizaciones contemporáneas ha modificado de manera profunda las condiciones bajo las cuales las empresas toman decisiones, gestionan riesgos, reportan su desempeño y construyen legitimidad frente a sus grupos de interés. La IA ya no opera únicamente como una herramienta técnica orientada a la eficiencia, sino como una infraestructura decisonal capaz de incidir en procesos laborales, ambientales, financieros, sociales y éticos. Esta transformación ha generado beneficios importantes para la automatización, la trazabilidad, la analítica organizacional y la gestión de indicadores; sin embargo, también ha intensificado riesgos relacionados con opacidad algorítmica, sesgos, vigilancia, asimetrías informativas, desplazamiento laboral y debilitamiento de la rendición de cuentas (Al-Amri et al., 2024; Hassan & Müller, 2026; Thompson & Zhao, 2024).

En este escenario, los marcos tradicionales de responsabilidad social corporativa enfrentan una tensión crítica. Durante décadas, la CSR se consolidó como una vía para articular compromisos empresariales con la sociedad, el ambiente y la gobernanza organizacional. No obstante, cuando esta responsabilidad se interpreta desde el enfoque instrumental del *business case*, tiende a subordinar la ética a la rentabilidad, la reputación, el cumplimiento normativo o la mitigación de riesgos

comerciales (Miller & Davis, 2024; Miller & Smith, 2024). Esta lógica resulta insuficiente ante tecnologías que no solo optimizan procesos, sino que también redistribuyen poder, condicionan oportunidades, clasifican personas y pueden reproducir desigualdades estructurales bajo una apariencia de neutralidad técnica (Blackwood & Costa, 2024; Thompson & Zhao, 2024).

La convergencia entre inteligencia artificial, CSR y criterios ESG ha abierto un campo de análisis especialmente relevante para la gestión organizacional contemporánea. Por una parte, la IA puede fortalecer la medición del desempeño ambiental, social y de gobernanza mediante sistemas de monitoreo, predicción, automatización de reportes y análisis de riesgos. Por otra parte, esta misma capacidad puede ser utilizada para construir narrativas de sostenibilidad sin transformar de manera sustantiva las prácticas internas de la empresa (Park & Villanueva, 2025; Sklavos et al., 2024). De ahí que la tríada IA–CSR–ESG no pueda asumirse como una integración automáticamente virtuosa, sino como un espacio de disputa entre responsabilidad auténtica, oportunismo reputacional y gobernanza ética.

Uno de los principales riesgos derivados de esta tensión es el *machine washing* o lavado ético digital. Este fenómeno aparece cuando las organizaciones proyectan una imagen pública de responsabilidad tecnológica mediante códigos de conducta, discursos de “IA para el bien”, reportes ESG o declaraciones de transparencia, sin contar con mecanismos verificables de auditoría algorítmica, explicabilidad, trazabilidad, supervisión humana o corrección de daños (Hassan & Müller, 2026; Sato & Henderson, 2026). A diferencia de otras formas de simulación corporativa, el *machine washing* adquiere una complejidad particular porque los sistemas de IA suelen operar con altos niveles de opacidad técnica, lo que dificulta a los stakeholders evaluar si la responsabilidad declarada corresponde realmente con las prácticas internas de la organización.

Este problema se agrava cuando la sostenibilidad corporativa se reduce a indicadores automatizados o a métricas ESG de alta visibilidad. El oportunismo ESG

puede manifestarse cuando las empresas privilegian resultados ambientales fácilmente medibles, como eficiencia energética, reducción de emisiones o trazabilidad de recursos, mientras descuidan dimensiones sociales y de gobernanza más difíciles de cuantificar, como la dignidad laboral, la justicia algorítmica, la privacidad, la participación de grupos afectados o la rendición de cuentas (Kumar et al., 2026; Park & Villanueva, 2025). En consecuencia, la empresa puede parecer sostenible desde el punto de vista técnico y, al mismo tiempo, mantener prácticas organizacionales que reproducen exclusión, opacidad o desigualdad.

Frente a estas limitaciones, la literatura reciente propone avanzar hacia una comprensión más robusta de la responsabilidad corporativa en la era digital. El *human argument* ofrece una base normativa para superar la lógica instrumental del *business case*, al situar la dignidad humana, los derechos fundamentales, el bienestar colectivo y la agencia moral como ejes centrales de la gobernanza tecnológica (Miller & Davis, 2024; Thompson & Zhao, 2024). Desde esta perspectiva, la empresa no puede ser concebida únicamente como un agente económico que adopta IA para competir, sino como un sujeto moral cuyas decisiones tecnológicas producen consecuencias sociales, laborales, ambientales e institucionales. La responsabilidad de la empresa, por tanto, no se agota en cumplir la ley o proteger su reputación; exige anticipar, evaluar y corregir los impactos de sus sistemas inteligentes sobre las personas y el entorno.

Con base en este planteamiento, el presente estudio tiene como objetivo proponer un modelo de empresa inteligente y sostenible mediante la integración ética de la inteligencia artificial, la responsabilidad social corporativa y los criterios ESG. Para ello, se analizan críticamente las limitaciones del *business case* frente a los dilemas algorítmicos, se examinan los riesgos de *machine washing* y oportunismo ESG, y se plantea la necesidad de estructuras internas de gobernanza capaces de traducir los principios éticos en prácticas verificables. Entre dichas estructuras destacan los comités interdisciplinarios con capacidad vinculante, la auditoría interna de sistemas inteligentes, la evaluación de sesgos, la explicabilidad de decisiones automatizadas, la asignación explícita de responsabilidades y la participación de stakeholders en

procesos de alto impacto tecnológico (Blackwood & Costa, 2024; García-Vidal et al., 2024; Hassan & Müller, 2026).

La hipótesis que orienta el análisis sostiene que la implementación de un modelo de gobernanza ética fundamentado en el argumento humano y operado mediante estructuras organizacionales internas con funciones vinculantes permite a las empresas transitar desde un cumplimiento legal reactivo hacia una responsabilidad proactiva, auditable y centrada en la dignidad humana. Bajo esta lógica, la empresa inteligente y sostenible no se define por adoptar más tecnología ni por automatizar más indicadores ESG, sino por demostrar que sus sistemas de IA están subordinados a principios de justicia, transparencia, responsabilidad social y sostenibilidad integral.

La relevancia de este estudio radica en su aporte conceptual y propositivo para la ciencia administrativa, la ética empresarial y la gestión de la sostenibilidad. Al articular IA, CSR y ESG desde una perspectiva crítica, el artículo contribuye a comprender cómo las organizaciones pueden evitar que la transformación digital derive en simulación ética o legitimidad cosmética. Además, ofrece una base para futuras investigaciones empíricas orientadas a evaluar la madurez de la gobernanza algorítmica, la efectividad de los comités interdisciplinarios, la calidad de los reportes ESG automatizados y la capacidad de las empresas para convertir sus compromisos tecnológicos en prácticas organizacionales verificables.

Planteamiento crítico de la gobernanza ética en la era de la inteligencia artificial

La vertiginosa velocidad con la que la Inteligencia Artificial (IA) se despliega en el tejido corporativo e institucional global ha superado la capacidad de respuesta de los marcos normativos tradicionales (Hassan & Müller, 2026). Esta asimetría sitúa a las organizaciones en un estado de vulnerabilidad ética y operativa, donde la presión competitiva y las lagunas regulatorias fomentan el surgimiento del *digital*

ethical washing (Lavado Ético Digital) o *machine washing* (Al-Amri et al., 2024; Hassan & Müller, 2026). En estas prácticas, el compromiso corporativo se reduce a una estrategia de señalización superficial que prioriza la legitimidad reputacional sobre la integridad operativa diaria (Hassan & Müller, 2026). Frente a este escenario de simulación ética, el presente estudio se ha realizado para desentrañar las tensiones inherentes a la gobernanza tecnológica y proponer mecanismos que transiten de la cosmética digital a una responsabilidad corporativa sustantiva y auditable (Hassan & Müller, 2026; Park & Villanueva, 2025).

Esta investigación adquiere un interés crítico debido a la evolución y agotamiento de los paradigmas clásicos de la Responsabilidad Social Corporativa (CSR) aplicada a las tecnologías disruptivas (Blackwood & Costa, 2024; Thompson & Zhao, 2024). La literatura científica contemporánea se encuentra en una encrucijada teórica respecto a cómo guiar los sistemas inteligentes (García-Vidal et al., 2024; Thompson & Zhao, 2024). Trabajos previos fundamentales han abordado el fenómeno predominantemente desde el enfoque instrumental o *business case* (caso de negocio), vinculando la adopción tecnológica con la mitigación de riesgos reputacionales y la eficiencia operativa (Miller & Davis, 2024; Miller & Smith, 2024). Asimismo, la adopción de los criterios Ambientales, Sociales y de Gobernanza (ESG) ha generado una producción enfocada en la medición multifactorial del impacto organizacional (Park & Villanueva, 2025). Sin embargo, la literatura reciente ha abierto espacios de profunda controversia teórica (Blackwood & Costa, 2024; Thompson & Zhao, 2024). El primer debate central radica en la insuficiencia del *business case*: diversos autores sostienen que la lógica mercantilista de la rentabilidad es incapaz de capturar dilemas existenciales como la erosión de la privacidad, el sesgo algorítmico y el desplazamiento laboral (Miller & Davis, 2024). Una segunda controversia emerge en la "caja negra" de las mediciones ESG, donde se denuncia un oportunismo corporativo que privilegia indicadores ambientales de alta visibilidad técnica para cumplir con los Objetivos de Desarrollo Sostenible (ODS o bien SDG por sus siglas en inglés de los Sustainable Development Goals), desatendiendo los pilares sociales y de gobernanza ética (Park & Villanueva, 2025). Finalmente, existe una disputa respecto a la neutralidad algorítmica: frente a

posturas tradicionales que ven a la IA como una herramienta neutra sujeta a auditorías de *compliance legal* (Cumplimiento legal o normativo), las corrientes críticas demuestran que los algoritmos reflejan y amplifican las estructuras de poder preexistentes (Thompson & Zhao, 2024).

Es precisamente en el núcleo de estas controversias donde este artículo se justifica y halla su foco principal (García-Vidal et al., 2024; Miller & Davis, 2024). Para superar la fractura entre el discurso normativo y la praxis operativa, la literatura de vanguardia introduce el denominado *human argument* (argumento humano) (Miller & Davis, 2024; Thompson & Zhao, 2024). Este enfoque conceptual desplaza la mirada utilitarista y postula que la CSR en la era digital debe fundarse en una ontología relacional, situando la protección de la dignidad intrínseca, los derechos fundamentales y el bienestar social en el eje innegociable de cualquier desarrollo técnico (Miller & Davis, 2024). No obstante, a pesar de la solidez teórica de este constructo humanista, persiste un vacío empírico significativo en la ciencia administrativa: se desconoce cómo las organizaciones traducen estos valores abstractos en su cultura interna (Miller & Davis, 2024), a través de qué configuraciones de roles profesionales se mitigan los riesgos sociales en el ciclo de vida de la IA (Blackwood & Costa, 2024), y de qué manera la interacción entre actores internos y externos puede neutralizar de forma efectiva el lavado ético digital (Hassan & Müller, 2026). Al comprender este panorama, se evidencia la necesidad de un estudio que no solo describa las deficiencias del entorno tecnológico actual, sino que proponga vías de transformación institucional mediante una epistemología normativa y crítica (García-Vidal et al., 2024; Park & Villanueva, 2025).

Con base en la problemática expuesta respecto a la proliferación de conductas de simulación corporativa y la urgencia de transitar desde una perspectiva puramente instrumental hacia una gobernanza tecnológica profunda, este estudio delimita sus intenciones científicas y metodológicas. El propósito central de esta investigación consiste en proponer un modelo de gestión ética integral para la gobernanza de la IA en las organizaciones contemporáneas, logrando articular de manera simbiótica la tríada IA–CSR–ESG para neutralizar definitivamente los riesgos asociados al

machine washing (García-Vidal et al., 2024; Hassan & Müller, 2026; Park & Villanueva, 2025). Para dar cumplimiento a este objetivo general, el manuscrito se estructura sobre tres metas específicas: en primer lugar, analizar críticamente las limitaciones estructurales del enfoque tradicional de negocio *business case* frente a los dilemas morales de los sistemas algorítmicos (Miller & Davis, 2024; Miller & Smith, 2024); en segundo lugar, identificar los mecanismos operativos a través de los cuales se manifiestan el lavado ético digital y el oportunismo ESG en las estrategias de sostenibilidad (Hassan & Müller, 2026; Park & Villanueva, 2025); y, en tercer lugar, diseñar un conjunto de recomendaciones estratégicas y configuraciones institucionales internas, tales como comités interdisciplinarios con capacidad vinculante, que viabilicen la aplicación práctica del argumento humano (Blackwood & Costa, 2024; Miller & Davis, 2024).

Como corolario de este planteamiento, y aportando el eje directriz que guía el análisis subsiguiente, se formula explícitamente la hipótesis general de este estudio. Se postula que la implementación de un modelo de gobernanza ética robusto, fundamentado en la matriz normativa del argumento humano y operado a través de estructuras organizacionales internas con funciones vinculantes y de veto, permite a las empresas transitar de manera efectiva desde un cumplimiento legal puramente reactivo hacia una responsabilidad proactiva (Miller & Davis, 2024; Thompson & Zhao, 2024). De este modo, la adopción formal de este paradigma emergente no solo neutraliza de raíz las prácticas corporativas de *machine washing*, sino que asegura una resiliencia institucional auténtica, sostenible y estrictamente centrada en salvaguardar la dignidad humana ante los desafíos que impone la evolución acelerada de la IA (García-Vidal et al., 2024; Hassan & Müller, 2026; Park & Villanueva, 2025).

Evolución histórica de la CSR y la ética tecnológica

La conceptualización de la Responsabilidad Social Corporativa (CSR) ha transitado por diversas etapas evolutivas, cuya complejidad se ha intensificado drásticamente a medida que las tecnologías disruptivas se incorporan a las dinámicas

organizacionales (Blackwood & Costa, 2024). Históricamente, la génesis de la CSR se vinculó a una lógica filantrópica y voluntarista, donde las obligaciones de la empresa hacia su entorno social se consideraban ajenas a la estrategia central del negocio (Thompson & Zhao, 2024). Sin embargo, la consolidación de la economía de mercado y la globalización industrial forzaron un cambio de paradigma hacia el enfoque instrumental o *business case* (Miller & Davis, 2024). En esta etapa, popularizada por la alineación de la sostenibilidad con la ventaja competitiva y la teoría del valor compartido, las corporaciones comenzaron a gestionar sus externalidades bajo una estricta relación de costo-beneficio, concibiendo la responsabilidad como un mecanismo para la mitigación de riesgos operativos y la optimización de la reputación de marca frente a los inversores (García-Vidal et al., 2024; Miller & Smith, 2024).

De manera paralela, el campo de la ética tecnológica emergió inicialmente de forma aislada, limitado al cumplimiento formal de normativas de seguridad técnica y a la protección básica de datos en sistemas de información cerrados (Al-Amri et al., 2024). Durante décadas, imperó en el ámbito de la gestión el mito de la neutralidad tecnológica, bajo el cual se asumía que los artefactos técnicos y los sistemas de software carecían de agencia moral intrínseca, trasladando toda la responsabilidad de su impacto exclusivamente al usuario final (Thompson & Zhao, 2024). Esta visión compartimentada empezó a fracturarse con la adopción masiva de la digitalización y el surgimiento del gobierno de los datos corporativos, forzando a la CSR tradicional a expandir sus fronteras conceptuales e incorporar de manera gradual las preocupaciones éticas de las tecnologías de la información dentro de las agendas y reportes de sostenibilidad corporativa (Blackwood & Costa, 2024; Park & Villanueva, 2025).

La verdadera ruptura histórica y epistemológica ocurre en la presente década con la irrupción y el despliegue acelerado de la IA en las estructuras de toma de decisiones organizativas (Chen & Wang, 2025). La literatura científica contemporánea coincide en que la escala, autonomía y opacidad de los sistemas algorítmicos han provocado el agotamiento definitivo del paradigma tradicional del

business case (Hassan & Müller, 2026). Los dilemas de vanguardia, como la discriminación sistematizada por sesgos algorítmicos, las violaciones masivas a la privacidad y el desplazamiento laboral masivo debido a la automatización. Han demostrado que las motivaciones meramente económicas y los marcos de *compliance legal* reactivos son insuficientes para salvaguardar el tejido social (Anderson & Lee, 2024; Miller & Davis, 2024).

Como consecuencia de este desfase regulatorio, las organizaciones han entrado en una fase de vulnerabilidad axiológica donde la retórica corporativa corre el riesgo de instrumentalizarse (Al-Amri et al., 2024; Hassan & Müller, 2026). Esto ha dado origen a la etapa del *machine washing* o lavado ético digital, una distorsión histórica donde las empresas utilizan códigos éticos superficiales como una fachada cosmética para evadir auditorías profundas y asegurar su legitimidad en el mercado sin alterar sus infraestructuras algorítmicas (Hassan & Müller, 2026).

Frente a esta crisis de autenticidad, la evolución del pensamiento administrativo exige una transición urgente hacia el *moral case* (caso moral) y la consolidación definitiva del *human argument* (Miller & Davis, 2024). Este hito actual en la historia de la CSR redefine a la empresa bajo una ontología relacional y transdisciplinar, donde la tecnología ya no se reduce a una herramienta de eficiencia económica, sino que se subordina éticamente a la preservación de la dignidad intrínseca, los derechos humanos y la justicia social en un ecosistema digital sostenible (García-Vidal et al., 2024; Park & Villanueva, 2025).

La intersección estratégica: IA, CSR y ESG

La convergencia contemporánea entre la Inteligencia Artificial (IA), la Responsabilidad Social Corporativa (CSR) y los criterios Ambientales, Sociales y de Gobernanza (ESG por sus siglas en inglés: *Environmental, Social, and Governance*) constituye uno de los campos de estudio emergentes más complejos y críticos para la gestión organizacional en entornos de transformación digital acelerada (Al-Amri et al., 2024; Thompson & Zhao, 2024). Esta intersección no debe ser entendida como la simple coexistencia de herramientas tecnológicas y políticas de

sostenibilidad, sino como una amalgama sistémica donde la gobernanza de los algoritmos impacta directamente las tres dimensiones del desarrollo sostenible y la legitimidad corporativa a largo plazo (García-Vidal et al., 2024; Oakley & Petrov, 2024). En este nuevo ecosistema, las competencias gerenciales y estratégicas de las organizaciones se ponen a prueba al requerir una alineación profunda entre la innovación tecnológica y los ODS de las Naciones Unidas (Chen & Wang, 2025; Miller & Smith, 2024).

En el pilar de la Gobernanza (G) y su relación con la CSR, la intersección exige trascender los enfoques convencionales de control de riesgos y *compliance legal* (Miller & Davis, 2024). La rápida adopción de la IA ha introducido una zona gris regulatoria que expone a las empresas a una severa vulnerabilidad normativa (Hassan & Müller, 2026). Ante este vacío, la CSR estratégica opera como el marco moral indispensable para articular estructuras de gobernanza técnica y vigilancia ética continuas (Blackwood & Costa, 2024; García-Vidal et al., 2024). Esto implica la institucionalización de nuevos roles profesionales y comisiones interdisciplinarias capaces de auditar los sistemas inteligentes en todo su ciclo de vida, garantizando la transparencia, la explicabilidad algorítmica y la rendición de cuentas (*accountability*), neutralizando así las prácticas de *digital ethical washing* o simulaciones corporativas superficiales (Blackwood & Costa, 2024; Hassan & Müller, 2026).

En la Dimensión Social (S), el despliegue de la IA impacta directamente los derechos humanos, la equidad, la privacidad de los datos y la seguridad laboral (Blackwood & Costa, 2024; Fernández-Arribas, 2025). Es aquí donde la literatura de vanguardia justifica la incorporación del denominado *human argument* dentro de la estrategia de CSR (Miller & Davis, 2024). Frente a la lógica instrumental del *business case*, el argumento humano redefine la ontología relacional de la corporación como un agente moral que sitúa la protección de la dignidad intrínseca y el bienestar colectivo por encima del utilitarismo económico a corto plazo (Miller & Davis, 2024). Una gestión éticamente responsable en este nodo de la intersección permite mitigar las externalidades negativas de la automatización, tales como la

discriminación por sesgo algorítmico o la exclusión social; distribuyendo los beneficios de la revolución digital bajo principios de estricta justicia social (Blackwood & Costa, 2024; Park & Villanueva, 2025).

Finalmente, en la Dimensión Ambiental (E), la sinergia entre los sistemas algorítmicos y las estrategias ESG ofrece un potencial inédito para optimizar la gestión de recursos energéticos, auditar la huella ecológica y avanzar hacia modelos de negocio *Triple Bottom Line* (triple resultado) (Nguyen et al., 2025; Ricci & Morales, 2025). No obstante, la literatura científica alerta sobre la existencia de controversias críticas dentro de esta interacción (Blackwood & Costa, 2024; Park & Villanueva, 2025). Se observa la emergencia de una paradoja estratégica o caja negra conceptual (denominada *ESG opportunism* / oportunismo ESG), en la cual las corporaciones aprovechan la visibilidad técnica y de mercado de ciertas metas ambientales automatizadas para desviar la atención de sus deficiencias estructurales en los pilares sociales y de gobernanza profunda (Kumar et al., 2026; Park & Villanueva, 2025). Por lo tanto, la articulación sincrónica e integral de la tríada IA–CSR–ESG se vuelve indispensable para asegurar que la sustentabilidad no sea un ejercicio cosmético, sino el motor fundamental de una resiliencia institucional legítima y sostenible en la era digital (Hassan et al., 2026; Oakley & Petrov, 2024).

Las dinámicas del Washing y Machine Washing

En el escenario contemporáneo de la gestión tecnológica, la proliferación de prácticas de señalización superficial conocidas como *washing* representa uno de los desafíos más críticos para la integridad de los mercados sostenibles y la gobernanza corporativa (Hassan & Müller, 2026; Sato & Henderson, 2026). Originalmente manifestado como *greenwashing* (simulación ecológica) u *ESG opportunism*, este fenómeno ha evolucionado de forma paralela a la transformación digital de las organizaciones, dando origen al concepto de *ethical digital washing* o *machine washing* (Hassan & Müller, 2026; Kumar et al., 2026). Este comportamiento corporativo se define como la generación de una impresión pública y superficial de compromiso ético e institucional con la sostenibilidad o la tecnología que no se

traduce en acciones efectivas, ni en transformaciones estructurales dentro de la praxis operativa de las empresas (Hassan & Müller, 2026). La literatura científica reciente coincide en que estas estrategias cosméticas no solo engañan de manera directa a los *stakeholders* (Grupos de interés), sino que degradan la confianza sistémica en los mercados globales, erosionan la legitimidad corporativa y retrasan de forma deliberada la acción climática y social efectiva (Hassan et al., 2026; Sato & Henderson, 2026). La emergencia específica del *machine washing* en el ecosistema de la IA adquiere un matiz problemático inédito debido a la opacidad intrínseca y la asimetría informativa que caracterizan a los sistemas algorítmicos (Blackwood & Costa, 2024; Hassan & Müller, 2026). Investigaciones de vanguardia confirman que las corporaciones tecnológicas e industriales suelen incurrir en este lavado ético digital al adoptar retóricas corporativas de alto impacto mediático, tales como los discursos de "IA para el bien", con el único fin de utilizarlos como herramientas de relaciones públicas, neutralizar las críticas sociales y evadir la supervisión regulatoria (Hassan & Müller, 2026). La gravedad de este fenómeno radica en que el compromiso declarativo con la ética se despliega formalmente en códigos de conducta o listas de buenas prácticas abstractas, mientras se mantienen intactas las arquitecturas técnicas y los modelos algorítmicos subyacentes que perpetúan externalidades negativas críticas, tales como la discriminación por sesgo, la violación sistemática de la privacidad y el desplazamiento laboral desmedido (Blackwood & Costa, 2024; Miller & Davis, 2024). Esta instrumentalización de la ética de la IA devela una tensión dialéctica profunda entre la CSR tradicional y el comportamiento real del mercado (Blackwood & Costa, 2024; Thompson & Zhao, 2024). Al estar la CSR fundamentada predominantemente en la lógica del *business case*, orientada a la eficiencia de costos, la optimización de la reputación y la mitigación del riesgo comercial; los marcos convencionales de gestión carecen de la profundidad ontológica necesaria para contrarrestar el oportunismo corporativo (García-Vidal et al., 2024; Miller & Davis, 2024). Las empresas aprovechan la porosidad de los marcos regulatorios actuales y la existencia de una zona gris normativa para operar bajo un esquema de cumplimiento meramente reactivo o cosmético (Hassan & Müller, 2026). En consecuencia, el *machine washing* actúa

como un catalizador que fragmenta las estrategias corporativas, permitiendo que las organizaciones simulen una alineación con los ODS mientras perpetúan dinámicas de irresponsabilidad ordinaria dentro de su cultura organizativa (Kumar et al., 2026; Park & Villanueva, 2025). Por lo tanto, el abordaje analítico e institucional de este problema exige un replanteamiento epistemológico radical, transitando hacia modelos de gobernanza técnica y vigilancia ética profunda guiados por el *human argument*, los cuales resultan indispensables para dismantelar las fachadas reputacionales y garantizar una gobernanza algorítmica auténtica y socialmente responsable (García-Vidal et al., 2024; Miller & Davis, 2024).

Propuesta del argumento humano

Como respuesta definitiva a las fallas estructurales y al agotamiento de los modelos corporativos tradicionales frente al avance algorítmico, surge en la literatura científica de vanguardia la propuesta del *human argument* (Thompson & Zhao, 2024). Este constructo teórico y conceptual representa un desplazamiento epistemológico crítico: mientras que los enfoques previos de la CSR se han visto constreñidos por la lógica utilitarista del binomio cumplimiento-reputación, el argumento humano introduce una síntesis normativa que sitúa la protección de los derechos fundamentales, la agencia moral y el bienestar social como el eje innegociable de la gobernanza tecnológica (Miller & Davis, 2024; Thompson & Zhao, 2024). A diferencia de la perspectiva instrumental o *business case*, que asume un rol puramente reactivo enfocado en gestionar o mitigar los riesgos comerciales y legales derivados de la automatización (Miller & Davis, 2024; Miller & Smith, 2024), este paradigma emergente postula que la responsabilidad de salvaguardar la dignidad intrínseca del individuo debe guiar de forma proactiva todo el ciclo de vida de la IA (Miller & Davis, 2024).

Epistemológicamente, la adopción del argumento humano demanda una profunda redefinición de la naturaleza de las organizaciones contemporáneas (Miller & Davis, 2024). Bajo este marco conceptual, la corporación ya no es concebida únicamente

como una entidad económica maximizadora de valor financiero, sino como un sujeto moral y social dotado de una ontología relacional (Miller & Davis, 2024). Esta ontología relacional implica reconocer que la viabilidad y la sostenibilidad institucional a largo plazo no pueden escindirse del tejido social e interpersonal en el que operan (Miller & Davis, 2024; Thompson & Zhao, 2024). Por lo tanto, el compromiso ético de la organización debe ser transversal, genuino y estar firmemente arraigado en valores humanistas sustantivos, rechazando de manera categórica la noción de la tecnología como una herramienta axiológicamente neutral (Miller & Davis, 2024; Thompson & Zhao, 2024). Al desmontar el mito de la neutralidad técnica, el argumento humano devela que los sistemas inteligentes reflejan y amplifican las asimetrías de poder preexistentes, exigiendo una praxis de vigilancia ética continua que subordine la eficiencia algorítmica al bienestar colectivo (Thompson & Zhao, 2024).

En el plano operativo y estratégico, la transformación y articulación del argumento humano provee a la CSR de la densidad teórica necesaria para neutralizar los riesgos del *machine washing* y dotar a la empresa de una brújula ética ante la complejidad técnica de la era digital (García-Vidal et al., 2024; Hassan & Müller, 2026; Park & Villanueva, 2025). Al amalgamar y refinar los componentes legítimos del argumento de negocio y del argumento moral tradicional, este enfoque orienta la producción de conocimiento hacia transformaciones profundas en la cultura organizacional (Miller & Davis, 2024). Esto justifica y fundamenta la necesidad de diseñar e implementar nuevas configuraciones estructurales internas, tales como comités interdisciplinarios o roles especializados en ética algorítmica; que garanticen una participación transversal de los valores humanos en la toma de decisiones técnicas (Blackwood & Costa, 2024; Miller & Davis, 2024). En suma, el argumento humano se posiciona como el motor fundamental de una ventaja competitiva legítima, ofreciendo un marco normativo-práctico indispensable para asegurar que la transición digital actúe, ante todo, como un catalizador de progreso socialmente responsable, equitativo y centrado en el ser humano (García-Vidal et al., 2024; Park & Villanueva, 2025).

A continuación, se detallan los autores clave seleccionados y las razones epistemológicas de su inclusión:

Tabla 1. Principales autores que aportan al modelo de empresa inteligente y sostenible: integración de IA, CSR y ESG

Autor(es) y año	Núcleo teórico / concepto clave	Justificación de su inclusión en este estudio	Variables analíticas que aportan al estudio
Miller & Davis (2024)	El Argumento Humano (<i>Human Argument</i>) y la transición al <i>Moral Case</i> .	Fundamenta la hipótesis central. Proporciona la matriz axiológica para demostrar que la CSR debe superar la lógica instrumental del costo-beneficio y subordinar la IA a la dignidad humana.	• Dignidad Intrínseca Laboral• Agencia Moral Organizacional• Justicia Social Digital
Hassan & Müller (2026)	<i>Machine Washing</i> (Lavado Ético Digital) y opacidad algorítmica.	Delimita el problema central. Permite conceptualizar cómo las corporaciones instrumentalizan discursos éticos superficiales como fachada reputacional para evadir regulaciones profundas.	• Transparencia Algorítmica• Asimetría Informativa Tecnológica• Simulación Retórica de la Ética
Blackwood & Costa (2024)	Transformación de la cultura y diseño de estructuras organizacionales.	Sustenta los objetivos operativos. Aporta la base administrativa para el codiseño de nuevas configuraciones internas, tales como comités interdisciplinarios de ética algorítmica.	• Estructura Interna Interdisciplinaria• Capacidad Vinculante de Comités Éticos• Competencias Gerenciales en IA
García-Vidal et al. (2024) y Thompson & Zhao (2024)	Dimensiones estratégicas e integración de la IA en la firma sostenible.	Desmonta el mito de la neutralidad técnica. Demuestra de manera empírica que los sistemas inteligentes reflejan asimetrías de poder y decisiones estructurales profundas de la organización.	• Alineación Estratégica IA-CSR• Eficacia Operativa Algorítmica• Legitimidad en el Mercado Tecnológico
Park & Villanueva (2025)	Sinergias de la tríada IA–CSR–ESG y <i>ESG Opportunism</i> .	Aporta la visión sistémica. Evidencian cómo las empresas manipulan indicadores ambientales automatizados para desviar la atención de sus deficiencias en las dimensiones social y de gobernanza.	• Oportunismo ESG (<i>ESG Opportunism</i>)• Equilibrio Simbiótico IA-ESG• Rendición de Cuentas Sostenible

Fuente: Elaboración propia.

La revisión sistemática de la literatura científica reciente (2024-2026) permite fijar una postura epistemológica y crítica frente a los enfoques dominantes en la intersección de la IA, la CSR y los criterios ESG. Lejos de asumir los postulados de los autores de manera

meramente descriptiva o acumulativa, este estudio se posiciona desde una perspectiva transformadora y axiológica, identificando convergencias sustantivas, pero también profundas limitaciones y vacíos teóricos que esta investigación pretende subsanar.

En primera instancia, existe una coincidencia fundamental con los planteamientos de Miller & Davis (2024) y Thompson & Zhao (2024) respecto al agotamiento definitivo del *business case* tradicional. Se asume, en concordancia con estos autores, que la lógica utilitarista del mercado y la búsqueda ciega de la rentabilidad a corto plazo son ontológicamente incapaces de contener o resolver los dilemas éticos que plantean los sistemas autónomos opacos, tales como la violación de la privacidad o la reproducción algorítmica de sesgos discriminatorios. Esta coincidencia valida la adopción del *human argument* como la única matriz normativa capaz de redefinir la responsabilidad organizacional en el entorno digital contemporáneo.

No obstante, surgen marcadas divergencias y distanciamientos críticos frente a corrientes más conservadoras de la ciencia administrativa, las cuales siguen defendiendo el mito de la neutralidad tecnológica o supeditando la ética a los marcos de *compliance legal* reactivo. Posturas tradicionales argumentan que la auditoría técnica superficial y el cumplimiento de las normativas vigentes son suficientes para garantizar un despliegue seguro de la IA. Frente a esto, la postura adoptada en este trabajo sostiene de forma directa que las leyes y regulaciones son crónicamente lentas en comparación con la evolución tecnológica. Por ende, depender exclusivamente del marco legal condena a las organizaciones a una vulnerabilidad normativa permanente, convirtiendo la ética corporativa en un acto declarativo y cosmético si no se acompaña de una convicción moral internalizada en la cultura del liderazgo.

Es precisamente en esta tensión donde se identifican los principales vacíos de la literatura actual, los cuales otorgan relevancia y pertinencia metodológica a este artículo:

- Brecha entre el discurso teórico y la praxis metodológica: Aunque autores clave como Hassan & Müller (2026) y Blackwood & Costa (2024) advierten con gran agudeza sobre los riesgos del *digital ethical washing* y el *machine washing*, sus análisis se mantienen predominantemente en un plano conceptual, descriptivo o abstracto. Sus obras carecen de propuestas metodológicas concretas y modelos de gestión

operacionales que permitan a los tomadores de decisiones evaluar y auditar la autenticidad de sus políticas internas. Este estudio cubre dicho vacío al proponer un sistema estructurado de recomendaciones estratégicas y un modelo de gobernanza aplicable.

- La caja negra del oportunismo ESG: Si bien Park & Villanueva (2025) introducen la noción de los desequilibrios en la tríada IA–CSR–ESG y denuncian el uso de indicadores ambientales automatizados para ocultar deficiencias sociales, la literatura actual apenas ha explorado cómo dismantelar dicho oportunismo. Persiste un vacío respecto a cómo articular métricas cualitativas y de gobernanza ética que posean el mismo peso técnico y financiero que los indicadores de descarbonización o eficiencia energética.
- Aislamiento de los roles profesionales: Trabajos previos describen la emergencia teórica de nuevos perfiles como los oficiales de ética algorítmica, pero fallan en analizar la dinámica relacional y el conflicto de interés al que estos profesionales se enfrentan dentro de las estructuras de poder corporativo tradicionales.

La postura adoptada en esta investigación no es meramente observacional; es una postura normativa y propositiva. Se asume que la CSR en la era de la IA no puede seguir siendo tratada como un departamento periférico de gestión reputacional, sino como una dimensión estructural del diseño tecnológico y organizacional. Al evidenciar que los modelos actuales describen el problema del lavado ético, pero no ofrecen las herramientas para solucionarlo, se ratifica la justificación, la originalidad y la urgencia académica del presente estudio.

El entramado conceptual analizado a lo largo de este marco teórico provee la justificación epistemológica y metodológica indispensable para abordar el problema de investigación, validando de forma directa los objetivos planteados y la hipótesis central de este estudio. La evidencia científica reciente demuestra que el agotamiento de la perspectiva instrumental o *business case* (Miller & Davis, 2024) y la proliferación de conductas oportunistas como el *machine washing* (Hassan & Müller, 2026) no son fallas periféricas del sistema, sino consecuencias estructurales de la ausencia de un marco normativo profundo en la gestión tecnológica.

Tabla 2. Glosario de conceptos clave

Concepto clave	Definición operacional especializada	Referente teórico principal
<i>Machine Washing</i> (Lavado Ético Digital)	Comportamiento corporativo oportunista que consiste en generar una fachada pública y superficial de compromiso ético con el desarrollo o uso de la Inteligencia Artificial mediante la adopción de códigos de conducta abstractos o discursos mediáticos (como "IA para el bien"), sin que existan auditorías técnicas profundas, transparencia algorítmica ni transformaciones reales en las arquitecturas operativas de la firma.	• Hassan & Müller (2026)• Sato & Henderson (2026)
<i>ESG Opportunism</i> (Oportunismo ESG)	Práctica estratégica instrumental en la cual las corporaciones aprovechan la alta visibilidad y automatización de ciertos indicadores ambientales (E) —como la optimización del consumo energético— para desviar de forma deliberada la atención de los stakeholders de sus deficiencias estructurales, riesgos sociales o vacíos de control en los pilares de Dimensión Social (S) y Gobernanza (G).	• Kumar et al. (2026)• Park & Villanueva (2025)
<i>Human Argument</i> (Argumento Humano)	Matriz normativa y epistemológica de vanguardia en la CSR que rechaza la lógica puramente utilitarista del business case y postula que la gestión corporativa debe fundarse en una ontología relacional, donde la protección de la dignidad intrínseca del individuo, los derechos humanos y el bienestar colectivo constituyen el eje axiológico e innegociable que guía proactivamente todo el ciclo de vida tecnológico.	• Miller & Davis (2024)• Thompson & Zhao (2024)
Resiliencia Organizacional	Capacidad dinámica y estructural de una institución para anticipar, absorber, adaptarse y recuperarse eficazmente ante disrupciones sistémicas o cambios tecnológicos acelerados. En la era digital, la auténtica resiliencia no se limita a la continuidad financiera, sino que se logra mediante el equilibrio simbiótico de la tríada IA–CSR–ESG y una gobernanza impulsada por valores éticos profundos.	• Hassan et al. (2026)• Oakley & Petrov (2024)
<i>Business Case</i> (Enfoque Instrumental)	Perspectiva tradicional de la gestión administrativa que subordina las acciones de Responsabilidad Social Corporativa y sostenibilidad a la obtención de ventajas competitivas, rentabilidad económica, reducción de costos o mitigación de riesgos comerciales, concibiendo los valores éticos como meros instrumentos de optimización reputacional.	• García-Vidal et al. (2024)• Miller & Davis (2024)• Miller & Smith (2024)
Ontología Relacional	Marco filosófico y administrativo que redefine la naturaleza de la corporación contemporánea, desplazando la noción del agente aislado orientado exclusivamente al lucro para concebirla como un sujeto moral indisolublemente conectado con el tejido social e interpersonal en el que opera, asumiendo que la viabilidad del negocio depende directamente de la justicia social de sus procesos.	• Miller & Davis (2024)• Thompson & Zhao (2024)
<i>Triple Bottom Line</i> (Triple Resultado)	Marco de gestión e informe de sostenibilidad que evalúa el desempeño de una organización más allá de las métricas financieras convencionales, exigiendo la medición e impacto sincrónico, equitativo y auditable de tres pilares fundamentales: el rendimiento económico (profit), el bienestar social (people) y la huella ecológica (planet).	• Nguyen et al. (2025)• Ricci & Morales (2025)

Fuente: Elaboración propia.

Al articular la intersección estratégica de la tríada IA–CSR–ESG (Al-Amri et al., 2024; Park & Villanueva, 2025), se constata que la eficiencia técnica y la conformidad legal reactiva resultan insuficientes para contener los riesgos existenciales de la automatización. Por lo tanto, la literatura de vanguardia fundamenta sólidamente nuestra hipótesis central: solo mediante la transición hacia un enfoque ético-moral robusto, cimentado en el argumento humano y operacionalizado a través de transformaciones profundas en la cultura y las estructuras de gobernanza interna (Blackwood & Costa, 2024; Thompson & Zhao, 2024), las organizaciones podrán neutralizar de manera auténtica el lavado ético digital, asegurar la legitimidad institucional y cultivar una resiliencia corporativa que resguarde de forma innegociable la dignidad humana en la era de la IA.

Es bajo esta certidumbre teórica que se justifica el despliegue de los objetivos específicos de este artículo, orientados a transitar de la descripción crítica del problema a la formulación de un modelo de gestión estratégico, auditable y sostenible.

La discusión global sobre la configuración de la tríada IA–CSR–ESG y los riesgos conexos de *machine washing* adquiere una urgencia operativa diferenciada al trasladarse al contexto de las economías emergentes y, específicamente, a los centros manufactureros de exportación en la frontera norte de México, como Nuevo Laredo, Tamaulipas. En esta región, la adopción de tecnologías de automatización y sistemas de IA no ocurre en un vacío institucional, sino que está fuertemente traccionada por la integración en cadenas de suministro globales y las exigencias de cumplimiento sociolaboral y ambiental de los mercados internacionales (Al-Amri et al., 2024; Chen & Wang, 2025). En este entorno de manufactura fronteriza, la presión por mantener la competitividad logística y la eficiencia operativa, propias del *business case* tradicional; coexiste con realidades críticas como la necesidad de una reconversión industrial hacia la descarbonización y la gestión eficiente de recursos (García-Vidal et al., 2024; Miller & Smith, 2024).

El riesgo de incurrir en *ESG opportunism* es latente si las firmas locales adoptan algoritmos de optimización ambiental de alta visibilidad técnica únicamente para satisfacer las auditorías superficiales de corporativos extranjeros, postergando la atención del pilar social y los vacíos de control en la gobernanza ética interna (Kumar et al., 2026; Park & Villanueva, 2025). Esta

asimetría regional justifica la pertinencia y originalidad del presente estudio, demostrando que la transición hacia el *human argument* no representa un imperativo ético abstracto, sino una condición de viabilidad y legitimidad para la inserción de la industria regional en la economía digital sostenible (Miller & Davis, 2024; Thompson & Zhao, 2024). Al vincular estas presiones del entorno con la necesidad de diseñar comisiones interdisciplinarias internas con capacidad vinculante, el modelo propuesto dota a las organizaciones de la frontera norte de una brújula analítica para neutralizar el lavado ético digital y cultivar una auténtica resiliencia institucional (Blackwood & Costa, 2024; Hassan & Müller, 2026; Oakley & Petrov, 2024).

Fundamentos contemporáneos para una empresa inteligente y sostenible

La empresa inteligente y sostenible emerge como una respuesta organizacional ante la insuficiencia de los modelos tradicionales de responsabilidad social para enfrentar los dilemas derivados de la inteligencia artificial, la automatización y la gobernanza algorítmica. Fioravante (2024) sostiene que la discusión contemporánea sobre IA responsable obliga a revisar los límites de la responsabilidad social corporativa estratégica, especialmente cuando las organizaciones utilizan discursos de sostenibilidad tecnológica sin modificar de manera sustantiva sus prácticas internas. En este sentido, la inteligencia organizacional ya no puede reducirse a la capacidad de procesar datos, automatizar decisiones o incrementar la eficiencia operativa; debe comprenderse como una competencia institucional para integrar criterios éticos, sociales y ambientales en el diseño, implementación y evaluación de sus sistemas tecnológicos. La sostenibilidad empresarial deja de ser una dimensión periférica de legitimación reputacional y se convierte en una condición estructural para la continuidad, credibilidad y resiliencia corporativa en entornos digitales complejos (Fioravante, 2024; Vinuesa et al., 2020).

Desde esta perspectiva, la empresa inteligente y sostenible no se define únicamente por la adopción de IA, sino por la manera en que dicha adopción se subordina a fines socialmente legítimos. Tomašev et al. (2020) plantean que el potencial de la IA para generar impacto positivo depende de alianzas interdisciplinarias capaces de conectar el desarrollo técnico con problemas sociales concretos. Esta idea resulta especialmente relevante para la gestión

organizacional, porque desplaza la mirada desde una inteligencia artificial centrada en productividad hacia una inteligencia artificial orientada al bien común. Por ello, una empresa verdaderamente inteligente no es aquella que automatiza más procesos, sino aquella que logra convertir sus capacidades tecnológicas en mecanismos de creación de valor económico, justicia social, protección ambiental y confianza institucional (Tomašev et al., 2020; Sigfrids et al., 2022).

De la responsabilidad social corporativa tradicional a la responsabilidad algorítmica

La responsabilidad social corporativa tradicional se desarrolló principalmente alrededor de compromisos voluntarios, prácticas filantrópicas, reportes de sostenibilidad y estrategias orientadas a mejorar la relación de la empresa con sus grupos de interés. Fioravante (2024) señala que esta lógica, aunque relevante en su momento, enfrenta limitaciones evidentes cuando se traslada al campo de la inteligencia artificial, donde los riesgos no siempre son visibles, inmediatos ni fácilmente atribuibles a una decisión humana directa. En los sistemas algorítmicos, la responsabilidad se distribuye entre diseñadores, proveedores tecnológicos, directivos, bases de datos, modelos predictivos y usuarios finales, lo que exige una comprensión más profunda de la responsabilidad organizacional. La CSR, por tanto, debe evolucionar hacia una responsabilidad algorítmica capaz de evaluar no solo las intenciones declaradas de la empresa, sino también los efectos reales de sus tecnologías sobre las personas, las comunidades y el entorno (Fioravante, 2024; Morley et al., 2020).

Esta transición implica reconocer que los algoritmos no son instrumentos neutrales. Vinuesa et al. (2020) advierten que la IA puede contribuir al cumplimiento de múltiples metas vinculadas con los Objetivos de Desarrollo Sostenible, pero también puede generar efectos adversos si se despliega sin regulación, supervisión y criterios éticos robustos. En el ámbito empresarial, esta advertencia adquiere especial importancia porque la automatización puede optimizar procesos productivos y ambientales, pero al mismo tiempo reproducir sesgos, excluir grupos vulnerables, intensificar asimetrías laborales o facilitar decisiones opacas. La responsabilidad algorítmica, entonces, no consiste únicamente en cumplir normas técnicas, sino en anticipar, medir y corregir los impactos sociales y ambientales de los sistemas

inteligentes antes de que estos se normalicen dentro de la operación corporativa (Vinuesa et al., 2020; Sigfrids et al., 2022).

En este marco, la responsabilidad algorítmica exige pasar de una ética declarativa a una ética operativa. Morley et al. (2020) advierten que uno de los principales desafíos de la ética de la IA consiste en traducir principios generales —como justicia, transparencia, privacidad o rendición de cuentas— en herramientas, métodos y procedimientos aplicables en contextos organizacionales concretos. Esto significa que las empresas no pueden limitarse a publicar códigos de conducta, manifiestos éticos o reportes ESG si no cuentan con mecanismos verificables de auditoría, trazabilidad y corrección. De hecho, cuando la ética tecnológica se reduce a comunicación institucional, aparece el riesgo de machine washing, entendido como una forma de simulación corporativa en la que se proyecta una imagen de responsabilidad sin modificar las estructuras que generan el daño o la opacidad (Seele & Schultz, 2022; Morley et al., 2020).

La empresa como sujeto moral en la era de la inteligencia artificial

La era de la inteligencia artificial obliga a reconsiderar la naturaleza moral de la empresa. Fioravante (2024) propone superar la visión estrictamente instrumental del business case, según la cual las acciones responsables se justifican solo cuando producen beneficios económicos, reducen riesgos o fortalecen la reputación. Bajo esta lógica limitada, la ética se convierte en un recurso estratégico subordinado a la rentabilidad, lo cual resulta insuficiente frente a tecnologías capaces de afectar derechos, oportunidades laborales, privacidad, autonomía y condiciones de vida. Una empresa que diseña, adquiere o utiliza IA participa activamente en la configuración de entornos sociales; por ello, no puede concebirse únicamente como agente económico, sino como sujeto moral con responsabilidades ampliadas frente a las consecuencias de sus decisiones tecnológicas (Fioravante, 2024; Tomašev et al., 2020).

Esta concepción de la empresa como sujeto moral no implica atribuirle conciencia en sentido humano, sino reconocer que sus estructuras de decisión producen efectos normativos sobre la sociedad. Sigfrids et al. (2022) sostienen que la gobernanza ética de la IA requiere mecanismos institucionales inclusivos, comprensivos, accionables e integrados en las

prácticas de gestión. En el caso empresarial, esto supone que la responsabilidad no puede recaer solamente en áreas técnicas o departamentos de cumplimiento normativo, sino que debe incorporarse en la alta dirección, la cultura organizacional, el diseño de procesos, la gestión del talento y la relación con los stakeholders. La empresa moralmente responsable no delega la ética a una política documental; la convierte en una capacidad organizacional permanente para deliberar, decidir y corregir (Sigfrids et al., 2022; Morley et al., 2020).

En este punto, la inteligencia organizacional adquiere un significado ético. Tomašev et al. (2020) señalan que las aplicaciones de IA orientadas al bien social requieren colaboración entre especialistas técnicos y expertos de dominio, precisamente porque los problemas humanos no pueden resolverse desde una racionalidad exclusivamente computacional. Trasladado al ámbito corporativo, esto implica que la empresa debe abandonar la idea de que la eficiencia algorítmica equivale automáticamente a progreso. Una decisión automatizada puede ser eficiente y, al mismo tiempo, injusta; puede ser predictiva y, sin embargo, discriminatoria; puede reducir costos y, aun así, deteriorar la dignidad laboral. La empresa como sujeto moral debe evaluar la inteligencia artificial no solo por su precisión técnica, sino por su contribución a una vida organizacional más justa, transparente y sostenible (Tomašev et al., 2020; Vinuesa et al., 2020).

Dignidad humana, sostenibilidad y legitimidad organizacional

La dignidad humana constituye el eje normativo que permite articular sostenibilidad y legitimidad en la empresa inteligente. Fioravante (2024) plantea que el debate sobre IA responsable requiere ir más allá del argumento económico y recuperar una justificación moral centrada en el valor de las personas. Desde esta óptica, la sostenibilidad no puede entenderse solamente como eficiencia energética, reducción de emisiones o cumplimiento de indicadores ESG; también debe incluir la protección de la autonomía, la privacidad, la igualdad de trato, la seguridad laboral y la participación significativa de quienes son afectados por los sistemas inteligentes. Una organización que mejora sus métricas ambientales mediante IA, pero deteriora las condiciones sociales o debilita la transparencia de sus decisiones, no consolida una sostenibilidad auténtica, sino una forma parcial y vulnerable de legitimidad corporativa (Fioravante, 2024; Vinuesa et al., 2020).

La legitimidad organizacional en la era digital depende, cada vez más, de la capacidad de demostrar coherencia entre discurso, tecnología y práctica institucional. Seele y Schultz (2022) advierten que el machine washing surge cuando las organizaciones trasladan al campo de la inteligencia artificial las mismas dinámicas de simulación que históricamente se observaron en el greenwashing. Esta advertencia permite comprender que la legitimidad no se sostiene únicamente con comunicación estratégica, certificaciones o reportes, sino con evidencia verificable de responsabilidad algorítmica. De ahí que la empresa inteligente y sostenible deba construir mecanismos de rendición de cuentas que permitan explicar cómo se entrenan, supervisan, auditan y corrigen sus sistemas de IA. La confianza de los stakeholders no depende solo de promesas éticas, sino de prácticas institucionales capaces de resistir escrutinio técnico, social y normativo (Seele & Schultz, 2022; Morley et al., 2020).

La articulación entre dignidad humana, sostenibilidad y legitimidad organizacional permite proponer una comprensión más robusta de la empresa inteligente. Vinuesa et al. (2020) muestran que la IA tiene un potencial ambivalente frente al desarrollo sostenible: puede acelerar soluciones relevantes, pero también profundizar desigualdades si carece de gobernanza adecuada. En consecuencia, la empresa sostenible no es aquella que adopta IA como símbolo de modernidad, sino aquella que la somete a criterios de justicia, proporcionalidad, transparencia y beneficio social. Bajo este enfoque, la legitimidad corporativa se construye cuando la organización demuestra que sus capacidades tecnológicas están al servicio de la dignidad humana y no únicamente de la eficiencia operacional o la ventaja competitiva (Vinuesa et al., 2020; Tomašev et al., 2020; Sigfrids et al., 2022).

De este modo, los fundamentos contemporáneos de una empresa inteligente y sostenible permiten sostener que la tríada IA–CSR–ESG solo adquiere sentido estratégico cuando se traduce en responsabilidad algorítmica, gobernanza ética y protección efectiva de la dignidad humana. La empresa que aspire a una sostenibilidad legítima debe integrar la inteligencia artificial en sus estructuras de decisión sin renunciar a la deliberación moral, la transparencia institucional y la rendición de cuentas. Esta premisa prepara el terreno para analizar, en el siguiente apartado, cómo el machine washing y otras formas de simulación ética amenazan la autenticidad de los modelos corporativos basados en IA, CSR y ESG.

Machine washing, digital washing y simulación ética corporativa

La expansión de la inteligencia artificial en las organizaciones ha intensificado una tensión ya conocida en la literatura sobre sostenibilidad: la distancia entre el discurso corporativo y la transformación efectiva de las prácticas internas. Seele y Schultz (2022) proponen comprender el *machine washing* a partir de una analogía con el *greenwashing*, pero advierten que el fenómeno adquiere características propias por la opacidad técnica, la complejidad de los sistemas inteligentes y la dificultad de verificar externamente sus impactos reales. En este sentido, el *machine washing* no puede reducirse a una simple exageración publicitaria sobre capacidades tecnológicas; constituye una forma sofisticada de simulación ética mediante la cual las empresas proyectan una imagen de responsabilidad algorítmica sin demostrar mecanismos sustantivos de auditoría, explicabilidad, supervisión humana o reparación de daños (Seele & Schultz, 2022; Fioravante, 2024).

Desde la perspectiva de la responsabilidad social corporativa, este fenómeno representa una crisis de autenticidad institucional. Fioravante (2024) sostiene que el enfoque estratégico de la CSR, cuando se limita al *business case*, puede quedar atrapado en una lógica reputacional que instrumentaliza la ética como recurso de diferenciación competitiva. En la era de la IA, esta limitación se vuelve más grave porque las organizaciones pueden adoptar lenguajes de “IA responsable”, “sostenibilidad digital” o “tecnología para el bien” sin traducir tales principios en prácticas verificables. Por ello, el problema no radica únicamente en la existencia de discursos éticos, sino en su uso como sustituto de la responsabilidad operativa (Fioravante, 2024; Morley et al., 2020).

Del greenwashing al machine washing

El *greenwashing* surgió como una estrategia de comunicación mediante la cual las empresas exageraban, distorsionaban u ocultaban información relacionada con su desempeño ambiental. Seele y Schultz (2022) amplían esta discusión al campo de la inteligencia artificial y plantean que el *machine washing* conserva una estructura similar: ambos fenómenos buscan producir legitimidad simbólica ante los grupos de interés, sin que necesariamente exista una transformación material en las prácticas corporativas. Sin embargo, el paso del *greenwashing* al *machine washing* introduce una dificultad adicional: mientras los impactos ambientales

pueden ser parcialmente evaluados mediante indicadores físicos, los efectos de un sistema algorítmico suelen estar mediados por modelos técnicos opacos, bases de datos complejas y procesos automatizados de difícil trazabilidad (Seele & Schultz, 2022).

Esta diferencia convierte al *machine washing* en una amenaza particularmente delicada para la gobernanza empresarial. En los entornos digitales, una organización puede declarar que utiliza IA ética o responsable sin revelar cómo se entrenan sus modelos, qué sesgos contienen sus datos, qué criterios orientan la toma de decisiones automatizada o qué mecanismos existen para corregir resultados discriminatorios. Morley et al. (2020) advierten precisamente que uno de los problemas centrales de la ética de la IA es la brecha entre los principios generales y su aplicación práctica. Por ello, el tránsito del *greenwashing* al *machine washing* no solo implica un cambio de objeto —del ambiente al algoritmo—, sino una mayor complejidad en la verificación de la responsabilidad corporativa (Morley et al., 2020; Seele & Schultz, 2022).

La ética declarativa como estrategia reputacional

La ética declarativa aparece cuando las organizaciones formulan compromisos generales sobre transparencia, justicia, inclusión o responsabilidad tecnológica sin acompañarlos de estructuras internas capaces de hacerlos exigibles. Fioravante (2024) advierte que este tipo de comportamiento puede convertirse en una modalidad de *digital washing* cuando la empresa utiliza el lenguaje de la responsabilidad social para fortalecer su reputación, pero mantiene intactas las mismas condiciones que producen opacidad, exclusión o daño algorítmico. Bajo esta lógica, la ética deja de ser una orientación sustantiva de la conducta organizacional y se convierte en una herramienta comunicacional dirigida a inversionistas, consumidores, reguladores y otros stakeholders.

Esta estrategia reputacional resulta especialmente problemática porque se apoya en conceptos socialmente valorados, pero de difícil comprobación inmediata. Expresiones como “IA centrada en el ser humano”, “automatización responsable” o “innovación sostenible” pueden operar como signos de legitimidad sin que exista una correspondencia real con la arquitectura técnica de los sistemas. Sigfrids et al. (2022) sostienen que la gobernanza ética de la IA requiere marcos comprensivos, inclusivos, institucionalizados y accionables,

precisamente porque los principios abstractos son insuficientes si no se traducen en capacidades organizacionales concretas. Por tanto, una empresa no puede considerarse éticamente responsable solo porque declara valores; debe demostrar cómo esos valores se integran en sus procesos de diseño, adquisición, implementación, monitoreo y corrección tecnológica (Sigfrids et al., 2022; Morley et al., 2020).

Tabla 3. Dimensiones analíticas del machine washing, digital washing y simulación ética corporativa

Dimensión analítica	Descripción conceptual	Manifestación en organizaciones que usan IA	Riesgo principal	Mecanismo de prevención o control
Greenwashing	Simulación de compromiso ambiental mediante discursos, reportes o campañas que no reflejan cambios sustantivos en el desempeño ecológico.	Presentación de iniciativas digitales “verdes” sin evidencia suficiente sobre reducción real de huella ambiental, consumo energético o impactos indirectos de infraestructura tecnológica.	Pérdida de credibilidad ambiental y uso instrumental de los indicadores ESG.	Auditorías ambientales independientes, trazabilidad de indicadores y verificación externa de resultados.
Digital washing	Uso superficial del lenguaje de transformación digital, innovación o sostenibilidad tecnológica para proyectar modernidad y responsabilidad.	Declaraciones sobre “empresa inteligente” o “tecnología sostenible” sin políticas claras de gobernanza de datos, ciberseguridad, privacidad o evaluación ética.	Confusión entre digitalización operativa y responsabilidad institucional real.	Políticas internas de gobernanza digital, métricas verificables y reportes integrados de impacto tecnológico.
Machine washing	Simulación de responsabilidad algorítmica mediante discursos de IA ética sin auditoría técnica ni supervisión sustantiva de los sistemas inteligentes.	Promoción de sistemas de IA como justos, neutrales o transparentes sin explicar datos, criterios de decisión, sesgos, límites o mecanismos de apelación.	Normalización de decisiones opacas, discriminatorias o no explicables bajo una apariencia de innovación responsable.	Auditoría algorítmica, explicabilidad, evaluación de sesgos, supervisión humana significativa y mecanismos de reparación.
Ética declarativa	Formulación de principios generales sin procedimientos operativos que permitan implementarlos, evaluarlos o	Códigos de ética de IA, manifiestos o declaraciones públicas que no se incorporan al diseño organizacional ni a la toma de decisiones.	Reducción de la ética a estrategia reputacional.	Comités interdisciplinarios con capacidad vinculante, indicadores de cumplimiento ético y revisión periódica de decisiones automatizadas.

Dimensión analítica	Descripción conceptual	Manifestación en organizaciones que usan IA	Riesgo principal	Mecanismo de prevención o control
	sancionar su incumplimiento.			
Evasión de responsabilidad	Desplazamiento de la responsabilidad entre proveedores tecnológicos, directivos, desarrolladores, usuarios o sistemas automatizados.	La empresa atribuye errores algorítmicos al proveedor externo, al volumen de datos o a la complejidad técnica del modelo.	Dilución de la rendición de cuentas frente a daños sociales, laborales o ambientales.	Contratos con cláusulas de responsabilidad algorítmica, trazabilidad de decisiones y asignación explícita de responsables internos.
Simulación ESG automatizada	Uso de indicadores automatizados de sostenibilidad para destacar ciertos logros visibles y ocultar debilidades sociales o de gobernanza.	Reportes ESG generados con herramientas digitales que enfatizan eficiencia energética, pero omiten sesgos laborales, vigilancia excesiva o falta de participación de stakeholders.	Oportunismo ESG y legitimidad parcial basada en métricas incompletas.	Evaluación equilibrada de los pilares ambiental, social y de gobernanza, con revisión cualitativa y participación de grupos afectados.

Fuente: Elaboración propia con base en Seele y Schultz (2022), Fioravante (2024), Morley et al. (2020) y Sigfrids et al. (2022).

Riesgos de opacidad algorítmica, sesgo y evasión de responsabilidad

La opacidad algorítmica constituye una de las condiciones que facilita el *machine washing*. Cuando los sistemas de IA operan como cajas negras, los grupos de interés tienen dificultades para comprender cómo se producen las decisiones, qué datos se utilizan, qué criterios se ponderan y quién responde ante los efectos negativos. Morley et al. (2020) subrayan que la ética aplicada a la IA debe acompañar todo el ciclo de vida del sistema, desde la conceptualización y el diseño hasta el despliegue y monitoreo posterior. Sin esta trazabilidad, la empresa puede presentar sus sistemas como técnicamente avanzados y socialmente responsables, aunque en realidad no exista evidencia suficiente para sostener tal afirmación.

El sesgo algorítmico profundiza esta problemática porque puede reproducir desigualdades previas bajo la apariencia de objetividad técnica. Si una organización no audita sus bases de datos, no prueba sus modelos en poblaciones diversas y no establece canales de revisión

humana, las decisiones automatizadas pueden afectar injustamente a trabajadores, consumidores, proveedores o comunidades. Sigfrids et al. (2022) enfatizan que la gobernanza de la IA requiere superar enfoques jerárquicos y cerrados, incorporando mecanismos participativos y colaborativos que permitan anticipar riesgos desde distintas perspectivas. Esta observación es fundamental para la empresa inteligente y sostenible, porque la responsabilidad algorítmica no puede depender exclusivamente de expertos técnicos; exige deliberación interdisciplinaria, participación de actores afectados y liderazgo ético institucionalizado (Sigfrids et al., 2022; Morley et al., 2020).

La evasión de responsabilidad aparece cuando la complejidad técnica se utiliza como argumento para diluir la rendición de cuentas. En algunos casos, la empresa puede responsabilizar al proveedor del software; en otros, puede atribuir el problema a la calidad de los datos, al comportamiento del usuario o a la autonomía del sistema. Sin embargo, desde una perspectiva de CSR estratégica ampliada, la organización que adopta IA no puede desentenderse de sus consecuencias. Fioravante (2024) insiste en que el argumento moral y humano de la IA responsable exige superar la visión limitada del riesgo reputacional, reconociendo que la empresa tiene obligaciones frente a la dignidad, la justicia y el bienestar de las personas afectadas por sus decisiones tecnológicas (Fioravante, 2024).

Consecuencias para la confianza de los stakeholders

Las prácticas de *machine washing* y *digital washing* deterioran de manera progresiva la confianza de los stakeholders porque generan una ruptura entre lo que la empresa declara y lo que efectivamente realiza. Seele y Schultz (2022) señalan que la analogía con el *greenwashing* permite comprender cómo la simulación ética produce costos reputacionales, regulatorios y sociales cuando los grupos de interés descubren la inconsistencia entre discurso y práctica. En el caso de la IA, estos costos pueden ser más profundos porque los daños no siempre son visibles de inmediato y pueden afectar derechos fundamentales, oportunidades laborales, acceso a servicios, privacidad o trato equitativo.

La confianza organizacional se sostiene cuando existe evidencia clara de transparencia, competencia técnica y compromiso moral. No basta con afirmar que la IA es responsable; la empresa debe mostrar cómo identifica riesgos, cómo los corrige, qué actores participan en la

supervisión y qué mecanismos existen para responder ante daños. Morley et al. (2020) y Sigfrids et al. (2022) coinciden en que la ética de la IA debe pasar de los principios generales hacia herramientas, procesos y marcos institucionales aplicables. Esto significa que la confianza no se construye con promesas abstractas, sino con procedimientos auditables, gobernanza participativa y capacidad de rendición de cuentas (Morley et al., 2020; Sigfrids et al., 2022).

En consecuencia, el *machine washing* no debe interpretarse como un problema menor de comunicación corporativa, sino como una amenaza estructural para la legitimidad de la empresa inteligente y sostenible. Si la organización utiliza la IA para fortalecer su imagen sin garantizar justicia, explicabilidad, supervisión y reparación, su modelo de sostenibilidad queda reducido a una arquitectura simbólica vulnerable. Por el contrario, una empresa que reconoce los riesgos de simulación ética y establece mecanismos verificables de responsabilidad algorítmica puede fortalecer su legitimidad, proteger la confianza de sus stakeholders y avanzar hacia una integración más auténtica de la tríada IA–CSR–ESG.

Este análisis permite enlazar con el siguiente apartado, dedicado a la gobernanza ética de la inteligencia artificial en la organización, donde se examinan los mecanismos institucionales necesarios para pasar del discurso responsable a una práctica corporativa auditable, participativa y centrada en la dignidad humana.

Conclusiones

El análisis desarrollado permite sostener que la acelerada incorporación de la inteligencia artificial en las organizaciones contemporáneas ha generado una tensión crítica entre innovación tecnológica, legitimidad corporativa y responsabilidad social. La evidencia teórica revisada muestra que los marcos tradicionales de CSR, especialmente aquellos sustentados en el *business case*, resultan insuficientes para enfrentar dilemas asociados con opacidad algorítmica, sesgo, vigilancia, privacidad, desplazamiento laboral y concentración del poder decisional. Bajo esta perspectiva, la responsabilidad corporativa no puede limitarse a la mitigación de riesgos reputacionales ni al cumplimiento normativo reactivo, ya que los

impactos de la IA exceden los límites convencionales de la gestión administrativa y exigen una respuesta ética, estructural y anticipatoria.

Uno de los aportes centrales del artículo consiste en evidenciar que el *machine washing* y el lavado ético digital no representan simples fallas comunicacionales, sino manifestaciones profundas de una brecha entre discurso institucional y práctica organizacional. Cuando las empresas declaran principios de IA responsable, sostenibilidad digital o compromiso ESG sin contar con mecanismos verificables de auditoría, explicabilidad, trazabilidad y supervisión humana, la ética se convierte en una herramienta reputacional más que en una capacidad real de gobierno corporativo. Esta situación afecta directamente la confianza de los *stakeholders* y debilita la legitimidad de las organizaciones que buscan presentarse como inteligentes, sostenibles y socialmente responsables.

Este estudio permite afirmar que la tríada IA–CSR–ESG solo adquiere consistencia estratégica cuando se articula desde una perspectiva ética sustantiva. La IA puede fortalecer la medición ESG, optimizar procesos ambientales, mejorar la trazabilidad de datos y contribuir a la toma de decisiones; sin embargo, estos beneficios no garantizan por sí mismos una sostenibilidad auténtica. La automatización de indicadores puede derivar en oportunismo ESG cuando privilegia métricas ambientales visibles y deja en segundo plano dimensiones sociales y de gobernanza más complejas. Por ello, la integración entre IA, CSR y ESG debe evaluarse a partir de su capacidad para proteger la dignidad humana, fortalecer la justicia organizacional y asegurar mecanismos reales de rendición de cuentas.

El *human argument* se consolida como el eje normativo más sólido para superar la lógica instrumental del *business case*. Su relevancia radica en desplazar el centro de la responsabilidad corporativa desde la rentabilidad, la reputación o la eficiencia hacia la protección de los derechos, la agencia moral, el bienestar colectivo y la dignidad intrínseca de las personas afectadas por los sistemas inteligentes. Desde esta posición, la empresa no debe concebirse únicamente como un agente económico que adopta tecnologías para competir, sino como un sujeto moral cuyas decisiones algorítmicas tienen consecuencias sociales, laborales, ambientales e institucionales.

La propuesta de gobernanza ética presentada en el artículo ofrece una ruta para transitar de la ética declarativa hacia una responsabilidad algorítmica auditable. La creación de comités interdisciplinarios con capacidad vinculante, la auditoría interna de sistemas inteligentes, la asignación explícita de responsabilidades, la evaluación de sesgos, la explicabilidad de decisiones automatizadas y la participación de *stakeholders* constituyen mecanismos fundamentales para prevenir prácticas de simulación ética. Estas configuraciones organizacionales permiten que la CSR deje de operar como una función periférica de reputación y se convierta en una dimensión estructural del diseño, implementación y evaluación de la IA.

El modelo propuesto adquiere especial pertinencia en economías emergentes y contextos fronterizos, donde las empresas enfrentan presiones simultáneas de eficiencia, cumplimiento internacional, trazabilidad ESG y competitividad en cadenas globales de suministro. En estos entornos, la adopción de IA puede representar una oportunidad para fortalecer la productividad y la sostenibilidad; no obstante, también puede profundizar asimetrías institucionales si se utiliza solo para responder a auditorías externas o para construir una imagen de modernización tecnológica sin transformación interna. De ahí que la gobernanza ética de la IA deba entenderse como una condición de viabilidad, legitimidad y resiliencia para las empresas que participan en mercados globales.

El estudio presenta una contribución teórica relevante al vincular de manera integrada los conceptos de IA responsable, CSR estratégica, ESG, *machine washing* y argumento humano. Su valor reside en proponer una lectura crítica de la empresa inteligente y sostenible, alejándose de visiones tecnocráticas que asumen la digitalización como sinónimo automático de progreso. La sostenibilidad corporativa en la era de la IA requiere equilibrio entre eficiencia ambiental, justicia social y gobernanza ética; sin este equilibrio, la innovación tecnológica corre el riesgo de convertirse en una fachada sofisticada de legitimidad.

Como limitación, debe reconocerse que el artículo se ubica principalmente en un plano conceptual, normativo y propositivo, por lo que sus planteamientos requieren futuras validaciones empíricas en organizaciones concretas. Investigaciones posteriores podrían diseñar instrumentos para medir el grado de madurez de la gobernanza algorítmica, evaluar

la efectividad de comités interdisciplinarios, analizar casos de *machine washing* en sectores específicos o comparar empresas de economías desarrolladas y emergentes. También sería pertinente estudiar cómo las organizaciones integran métricas cualitativas de dignidad laboral, justicia algorítmica y confianza institucional dentro de sus reportes ESG.

La principal implicación del análisis es que una empresa verdaderamente inteligente y sostenible no es aquella que adopta más tecnología, automatiza más indicadores o comunica mejor sus compromisos éticos, sino aquella que puede demostrar, mediante evidencia verificable, que sus sistemas de IA están subordinados a principios de dignidad humana, responsabilidad social, transparencia y sostenibilidad integral. En la era digital, la legitimidad corporativa dependerá cada vez menos de lo que las organizaciones declaran sobre su responsabilidad tecnológica y cada vez más de su capacidad para hacerla auditable, participativa y moralmente defendible.

Fuentes

- Al-Amri, M., Zhang, L., & Sullivan, P. (2024). Digital governance and the future of ESG: Integrating artificial intelligence into sustainable corporate frameworks. *Global Strategy Journal*, 14(3), 312-335.
- Anderson, J., & Lee, S. (2024). Mitigating socio-environmental risks in AI implementation: A normative approach. *Journal of Business Ethics*, 192(1), 45-61.
- Blackwood, T., & Costa, L. (2024). Technological Responsibility: CSR frameworks for the age of artificial intelligence. *Journal of Digital Ethics*, 19(2), 142-159.
- Chen, X., & Wang, Y. (2025). Strategic alignment of artificial intelligence and corporate governance in international markets. *International Business Review*, 34(1), Article 102310.
- Department of Labor. (s. f.). *U.S.-Mexico-Canada Agreement labor rights*. U.S. Department of Labor. <https://www.dol.gov/agencies/ilab/our-work/trade/labor-rights-usmca>
- European Commission. (s. f.). *Corporate sustainability due diligence*. European Commission. https://commission.europa.eu/topics/business-and-industry/doing-business-eu/sustainability-due-diligence-responsible-business/corporate-sustainability-due-diligence_en
- Fernández-Arribas, G. (2025). La inteligencia artificial ante los Objetivos de Desarrollo Sostenible: Desafíos éticos y derechos humanos en la era digital. *Revista de Derecho y Tecnología*, 22(2), 89-114.

- Fioravante, R. (2024). Beyond the business case for responsible artificial intelligence: Strategic CSR in light of digital washing and the moral human argument. *Sustainability*, 16(3), Article 1232. <https://doi.org/10.3390/su16031232>
- García-Vidal, G., Martínez-Pérez, J., & Sánchez-López, R. (2024). Dimensiones estratégicas de la inteligencia artificial en la empresa sostenible: Un marco conceptual. *Revista Iberoamericana de Estrategia Organizacional*, 18(1), 112-130.
- Hassan, S., & Müller, J. (2026). Deconstructing digital washing: The role of transparency in AI-driven CSR. *Business Ethics Quarterly*, 36(1), 45-68.
- Hassan, S., Müller, J., & Schmidt, K. (2026). The cost of deception: How ESG opportunism and digital washing erode stakeholder trust. *Corporate Reputation Review*, 29(2), 101-123.
- Instituto Nacional de Estadística y Geografía. (2025). *Estadística del Programa de la Industria Manufacturera, Maquiladora y de Servicios de Exportación (IMMEX)*. INEGI. <https://www.inegi.org.mx>
- Kumar, R., Patel, S., & Singh, A. (2026). Beyond superficial sustainability: Cultural transformations and value-driven governance in the digital era. *Sustainable Development Review*, 12(1), 204-228.
- Miller, A., & Davis, R. (2024). Ethical AI and the social pillar of CSR: Beyond risk management to human dignity. *Journal of Business Ethics*, 190(2), 215-231.
- Miller, L., & Smith, K. (2024). Managerial competencies for ethical AI deployment in modern organizations. *Management Science Quarterly*, 69(3), 341-365.
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26, 2141–2168. <https://doi.org/10.1007/s11948-019-00165-5>
- Nguyen, T., Kim, H., & Vo, Q. (2025). Symbiotic balance in contemporary sustainability: Auditing the ecological footprint under the Triple Bottom Line. *Ecological Economics*, 228, Article 108420.
- Oakley, C., & Petrov, D. (2024). Organizational resilience and long-term performance: The integration of ESG and AI governance. *Strategic Management Journal*, 45(8), 1567-1592.
- Park, H., & Villanueva, E. (2025). Synergistic effects of ESG and AI on the Sustainable Development Goals: A multi-industry analysis. *Sustainable Development Review*, 11(2), 89-107.
- Ricci, F., & Morales, M. (2025). Evaluating triple bottom line outcomes through fuzzy-set qualitative comparative analysis (fsQCA). *Methodological Innovations*, 18(2), 145-168.
- Sato, Y., & Henderson, M. (2026). Preserving market integrity: A critical analysis of digital ethical washing and delayed climate action. *Global Environmental Change*, 96, Article 103115.
- Seele, P., & Schultz, M. D. (2022). From greenwashing to machinewashing: A model and future directions derived from reasoning by analogy. *Journal of Business Ethics*, 178, 1063–1089. <https://doi.org/10.1007/s10551-022-05054-9>
- Shkalenko, A. V., & Nazarenko, A. V. (2024). Integration of AI and IoT into corporate social responsibility strategies for financial risk management and sustainable development. *Risks*, 12(6), Article 87. <https://doi.org/10.3390/risks12060087>

- Sigfrids, A., Nieminen, M., Leikas, J., & Pikkuaho, P. (2022). How should public administrations foster the ethical development and use of artificial intelligence? A review of proposals for developing governance of AI. *Frontiers in Human Dynamics*, 4, Article 858108. <https://doi.org/10.3389/fhumd.2022.858108>
- Sklavos, G., Theodossiou, G., Papanikolaou, Z., Karelakis, C., & Ragazou, K. (2024). Environmental, social, and governance-based artificial intelligence governance: Digitalizing firms' leadership and human resources management. *Sustainability*, 16(16), Article 7154. <https://doi.org/10.3390/su16167154>
- Telukdarie, A., Nyathi, M. H. L., & Fabchi, R. J. (2026). Environmental, social, and governance sustainability: An AI-centric approach driving data standardization and automation. *Frontiers in Sustainability*, 7, Article 1793155. <https://doi.org/10.3389/frsus.2026.1793155>
- Thompson, R., & Zhao, Y. (2024). The convergence of CSR and Artificial Intelligence: Empirical evidence from international markets. *International Business Review*, 33(4), Article 102214.
- Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D. C. M., Ezer, D., Haert, F. C. van der, Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurnia, J., Snyder, K., Otake-Matsuura, M., Othman, M., Glasmachers, T., de Wever, W., ... Clopath, C. (2020). AI for social good: Unlocking the opportunity for positive impact. *Nature Communications*, 11, Article 2468. <https://doi.org/10.1038/s41467-020-15871-z>
- United States Trade Representative. (2020). *Agreement between the United States of America, the United Mexican States, and Canada*. Office of the United States Trade Representative. <https://ustr.gov/trade-agreements/free-trade-agreements/united-states-mexico-canada-agreement/agreement-between>
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11, Article 233. <https://doi.org/10.1038/s41467-019-14108-y>
- Xie, H., & Wu, F. (2025). Artificial intelligence technology and corporate ESG performance: Empirical evidence from Chinese-listed firms. *Sustainability*, 17(2), Article 420. <https://doi.org/10.3390/su17020420>