

ARTÍCULO

Inteligencia artificial, desempeño ESG y creación de valor: un modelo teórico basado en capacidades organizacionales

Inteligência artificial, desempenho ESG e criação de valor: um modelo teórico baseado em capacidades organizacionais

Artificial intelligence, ESG performance, and value creation: a theoretical model based on organizational capabilities

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
02/04/2024	18/10/2024	06/12/2024	en proceso

RESUMEN

La inteligencia artificial puede ampliar la capacidad de previsión, seguimiento y eficiencia de las organizaciones, pero su adopción no garantiza por sí sola la mejora del desempeño ambiental, social y de gobernanza ni la creación sostenible de valor. Este artículo tiene como objetivo desarrollar un modelo teórico que explique cómo las capacidades organizacionales condicionan esta relación. Se adopta un enfoque teórico-integrador que articula la teoría de las capacidades dinámicas, la literatura sobre gobernanza de la inteligencia artificial y ética algorítmica, y los estudios consolidados sobre desempeño ESG y creación de valor. Se proponen cuatro capacidades mediadoras: integración y calidad de los datos, aprendizaje organizacional, gobernanza responsable y conversión de información en decisiones estratégicas sostenibles. El modelo identifica también cinco mecanismos de destrucción de valor asociados al uso de la inteligencia artificial: opacidad decisoria, sesgos y discriminación sistémica, consumo intensivo de recursos computacionales, baja responsabilidad y

desalineación estratégica. Se discuten además la orquestación y la complementariedad entre las capacidades, las trayectorias de madurez organizacional, los riesgos vinculados a la dependencia de proveedores tecnológicos y las implicaciones para la gobernanza corporativa. Se concluye que los beneficios de la inteligencia artificial para la agenda ESG dependen de la capacidad organizacional de combinar eficiencia, legitimidad, control de riesgos y resultados socioambientales, lo que distingue la simple adquisición de tecnología de la competencia efectiva para utilizarla de manera responsable y sostenible.

Palabras clave: inteligencia artificial; ESG; capacidades organizacionales; creación de valor; gobernanza.

RESUMO

A inteligência artificial pode ampliar as capacidades de previsão, monitoramento e eficiência das organizações, mas sua adoção, por si só, não garante a melhoria do desempenho ambiental, social e de governança nem a criação sustentável de valor. Este artigo desenvolve um modelo teórico que explica como as capacidades organizacionais condicionam essa relação. Adota-se uma abordagem teórico-integradora que articula a teoria das capacidades dinâmicas, a literatura sobre governança da inteligência artificial e ética algorítmica, e os estudos consolidados sobre desempenho ESG e criação de valor. Propõem-se quatro capacidades mediadoras: integração e qualidade dos dados, aprendizagem organizacional, governança responsável e conversão de informação em decisões estratégicas sustentáveis. O modelo identifica também cinco mecanismos de destruição de valor associados ao uso da inteligência artificial: opacidade decisória, vieses e discriminação sistêmica, consumo intensivo de recursos computacionais, baixa responsabilização e desalinhamento estratégico. O artigo discute ainda a orquestração e a complementaridade entre as capacidades, as trajetórias de maturidade organizacional, os riscos associados à dependência de fornecedores tecnológicos e as implicações para a governança corporativa. Conclui-se que os benefícios da inteligência artificial para a agenda ESG dependem da capacidade organizacional de combinar eficiência, legitimidade, controle de riscos e resultados socioambientais, distinguindo a mera aquisição de tecnologia da competência efetiva para utilizá-la de forma responsável e sustentável.

Palavras-chave: inteligência artificial; ESG; capacidades organizacionais; criação de valor; governança.

ABSTRACT

Artificial intelligence can expand forecasting, monitoring, and efficiency capabilities, but its adoption alone does not guarantee improved environmental, social, and governance performance or sustainable value creation. This article develops a theoretical model explaining how organizational capabilities condition this relationship. It adopts a theoretical-integrative approach that combines dynamic capabilities theory, the literature on artificial intelligence governance and algorithmic ethics, and established research on ESG performance and value creation. Four mediating capabilities are proposed: data integration and quality, organizational learning, responsible governance, and the conversion of information into sustainable strategic decisions. The model also identifies five mechanisms of value destruction associated with artificial intelligence: decision opacity, bias and systemic discrimination, intensive computational resource consumption, weak accountability, and strategic misalignment. The article further discusses the orchestration and complementarity among capabilities, organizational maturity trajectories, risks associated with dependence on technology vendors, and implications for corporate governance. It concludes that the benefits of artificial intelligence for the ESG agenda depend on an organization's ability to combine efficiency, legitimacy, risk control, and social and environmental outcomes, distinguishing mere technology acquisition from the effective capability to use it responsibly and sustainably.

Keywords: artificial intelligence; ESG; organizational capabilities; value creation; governance.

INTRODUCCIÓN

La adopción corporativa de tecnologías basadas en inteligencia artificial ha venido remodelando, a un ritmo cada vez más acelerado, los fundamentos de la estrategia competitiva, la arquitectura organizacional y los modelos de negocio a escala global. Al mismo tiempo, las presiones institucionales, regulatorias y de mercado asociadas a la agenda ESG (ambiental, social y de gobernanza) exigen que las empresas demuestren un compromiso genuino con la sostenibilidad ambiental, la equidad social y la integridad de la gobernanza, más allá de la retórica declaratoria presente en los informes institucionales. Este doble movimiento, marcado por la intensa digitalización de las operaciones y por demandas crecientes de resultados sociales y

ambientales medibles, ha llevado a gestores e investigadores a suponer, a menudo de forma acrítica, que la incorporación de herramientas algorítmicas avanzadas produce, casi automáticamente, ganancias en el desempeño ESG y en la creación de valor corporativo.

La evidencia disponible no respalda simplemente este supuesto. Adquirir sistemas de inteligencia artificial, contratar consultorías especializadas o implementar modelos aislados de aprendizaje automático no garantiza, por sí solo, mejoras consistentes en los indicadores ambientales, sociales y de gobernanza de una organización. Por el contrario, una parte considerable de la literatura crítica sobre tecnología y sociedad documenta externalidades significativas asociadas a estos sistemas: el alto consumo energético de los centros de datos (Strubell et al., 2019), la reproducción de sesgos discriminatorios en decisiones automatizadas (Buolamwini & Geburu, 2018), la opacidad en los procesos decisorios y la erosión de la legitimidad organizacional ante diversos grupos de interés (Suchman, 1995; Whittaker, 2021). La relación entre la adopción de la inteligencia artificial y la creación sostenible de valor no es, por lo tanto, ni directa ni automática. Está mediada por rutinas gerenciales, activos intangibles y capacidades organizacionales específicas que determinan si la tecnología se convertirá en una ventaja competitiva legítima o en una fuente adicional de riesgo.

La pregunta que guía este ensayo teórico puede formularse de la siguiente manera: ¿qué capacidades organizacionales median la relación entre la adopción de la inteligencia artificial y el desempeño ESG, condicionando la creación o destrucción neta de valor corporativo? Se sostiene a lo largo del texto que convertir el potencial técnico de la inteligencia artificial en resultados sostenibles depende del desarrollo deliberado de competencias internas orientadas a la integración y calidad de los datos, el aprendizaje organizacional continuo, la gobernanza algorítmica responsable y la capacidad de traducir los análisis técnicos en decisiones estratégicas orientadas simultáneamente hacia el valor económico y el valor social y ambiental.

El propósito de este artículo es proponer un modelo teórico que explique cómo estas capacidades organizacionales condicionan la relación entre inteligencia artificial, desempeño ESG y creación de valor. Para ello, el texto se estructura en ocho secciones, además de esta introducción. La segunda sección presenta el marco teórico, articulando la teoría de las capacidades dinámicas, la literatura sobre gobernanza de la inteligencia artificial y los estudios sobre desempeño ESG. La tercera sección describe el enfoque metodológico adoptado, de naturaleza teórico-integradora. La cuarta sección

expone el modelo propuesto y sus cuatro capacidades mediadoras, discutiendo además su orquestación, la madurez organizacional necesaria para sostenerlas y los riesgos asociados a la dependencia de proveedores tecnológicos. La quinta sección examina los mecanismos de destrucción de valor vinculados al uso de la inteligencia artificial en contextos ESG. La sexta sección discute las implicaciones gerenciales y estratégicas, incluyendo la arquitectura de gobernanza y la gestión del ciclo de vida de los sistemas. La séptima sección presenta proposiciones teóricas y una agenda de investigación. Finalmente, la octava sección ofrece las consideraciones finales.

Cabe hacer una observación inicial. Este artículo no pretende agotar la discusión sobre la inteligencia artificial en las organizaciones, ni propone un modelo definitivo. Su propósito es más modesto y, al mismo tiempo, más específico: ofrecer un marco conceptual que ayude a investigadores y gestores a distinguir entre el uso instrumental de la tecnología y la construcción de competencias organizacionales genuinamente capaces de sostener resultados ambientales, sociales y de gobernanza consistentes a lo largo del tiempo.

MARCO TEÓRICO

El marco teórico de este estudio articula tres cuerpos de literatura que, aunque a menudo se tratan por separado, mantienen estrechas relaciones conceptuales cuando se aplican al problema de la adopción de la inteligencia artificial en contextos ESG: la teoría de las capacidades dinámicas, los estudios sobre gobernanza de la inteligencia artificial y ética algorítmica, y la literatura sobre desempeño ESG y creación de valor. Esta articulación parte de una observación sencilla: los modelos que tratan la tecnología como un factor de productividad exógeno y neutral tienden a subestimar el papel de las decisiones organizacionales en la determinación de sus efectos. Los sistemas de inteligencia artificial son artefactos sociotécnicos cuyos resultados dependen tanto de sus propiedades técnicas como de las rutinas, los valores y las estructuras decisorias de las organizaciones que los implementan (Orlikowski, 2007). Esta premisa orienta la construcción del modelo presentado en la cuarta sección.

Capacidades dinámicas y ventaja competitiva en la economía digital

La Visión Basada en Recursos sostiene que la ventaja competitiva de una empresa proviene de la posesión de recursos valiosos, raros, difíciles de imitar y organizacionalmente explotados (Barney, 1991). Sin embargo, esta

perspectiva presenta limitaciones importantes en entornos de cambio tecnológico acelerado, en los que la mera posesión de recursos no garantiza un desempeño superior sostenido. La teoría de las capacidades dinámicas, desarrollada por Teece, Pisano y Shuen (1997) y posteriormente perfeccionada por Teece (2007), aborda esta limitación al desplazar el foco de la posesión estática de recursos hacia la capacidad organizacional de integrarlos, construirlos y reconfigurarlos en respuesta a los cambios del entorno competitivo. Según Teece (2007), esta capacidad se desdobra en tres microfundamentos: la capacidad de detección (sensing), orientada a identificar oportunidades y amenazas; la capacidad de aprovechamiento (seizing), relacionada con la movilización de recursos para capturar el valor identificado; y la capacidad de transformación (transforming), asociada a la reconfiguración continua de activos, estructuras y rutinas organizacionales.

Eisenhardt y Martin (2000) contribuyen a esta discusión al señalar que las capacidades dinámicas no constituyen, por sí mismas, una fuente directa de ventaja competitiva sostenible; funcionan más bien como procesos organizacionales que permiten a la empresa reconfigurar su base de recursos de manera coherente con los cambios del entorno. Esta distinción es relevante para el argumento de este artículo. Los recursos tecnológicos avanzados, incluidos los sistemas de inteligencia artificial, se originan esencialmente como herramientas de eficiencia operativa. Su transformación en un diferencial competitivo duradero depende del desarrollo de rutinas organizacionales de orden superior, capaces de absorber la tecnología, adaptarla a procesos decisorios específicos y realinearla con las demandas de reguladores, inversores y otros grupos de interés.

Al aplicar esta lógica al contexto de la inteligencia artificial, puede sostenerse que la tecnología por sí sola no constituye una fuente de ventaja competitiva sostenible, ni de un desempeño ESG superior. Los sistemas de aprendizaje automático, las plataformas analíticas y las herramientas de automatización están hoy ampliamente disponibles en el mercado, a menudo como servicios en la nube ofrecidos por un pequeño número de grandes proveedores. Esta disponibilidad reduce el potencial de diferenciación asociado a la mera adquisición de la tecnología (Wamba et al., 2021). El diferencial competitivo se desplaza así hacia las rutinas organizacionales que permiten a una empresa utilizar estos recursos de manera coherente con su estrategia y con las expectativas de sus grupos de interés, un argumento que sustenta la propuesta de las capacidades mediadoras discutidas en la cuarta sección de este artículo.

Cabe señalar que este desplazamiento del foco, de la posesión de recursos hacia la capacidad de orquestarlos, tiene consecuencias prácticas para la forma en que los consejos de administración y los inversores evalúan los proyectos corporativos de inteligencia artificial. Los anuncios de inversión tecnológica tienden a ser recibidos en los mercados de capitales como una señal positiva, asociada a expectativas de ganancias de eficiencia e innovación. La teoría de las capacidades dinámicas sugiere cautela respecto a esta lectura: el valor de la inversión depende menos del monto destinado a hardware, software o licenciamiento de modelos, y más de la existencia previa, o del desarrollo paralelo, de rutinas organizacionales capaces de absorber esta tecnología y reconfigurar los procesos a su alrededor. Los inversores y los miembros de los consejos que evalúan únicamente el volumen de recursos destinados a la inteligencia artificial, sin examinar la madurez organizacional subyacente, corren el riesgo de sobrestimar sistemáticamente el potencial de creación de valor de estas inversiones.

Gobernanza de la inteligencia artificial, ética algorítmica y legitimidad organizacional

La gobernanza corporativa tradicional fue concebida en gran medida para abordar los problemas de agencia entre accionistas y ejecutivos, con énfasis en mecanismos de control financiero y contable. La presencia creciente de sistemas algorítmicos en los procesos decisorios corporativos amplía el alcance de esta gobernanza, exigiendo mecanismos capaces de supervisar decisiones automatizadas que a menudo son opacas y difíciles de auditar. La literatura sobre ética y gobernanza de la inteligencia artificial se ha consolidado como un campo relevante para la gestión estratégica, reuniendo contribuciones sobre principios éticos aplicables a sistemas autónomos, explicabilidad técnica y mecanismos de responsabilización (Floridi et al., 2018; Jobin, Ienca & Vayena, 2019).

Jobin, Ienca y Vayena (2019), en una revisión de documentos orientadores sobre ética de la inteligencia artificial publicados por gobiernos, empresas y organizaciones internacionales, identifican una convergencia en torno a principios como transparencia, equidad, no maleficencia, responsabilización y privacidad, aunque también señalan divergencias significativas respecto a cómo deben operacionalizarse estos principios. Este hallazgo es importante para el argumento de este artículo: la existencia de principios declarados no equivale a la existencia de capacidades organizacionales capaces de aplicarlos de manera consistente. Floridi et al. (2018) refuerzan este punto al proponer un marco ético para los sistemas de inteligencia artificial que articula beneficencia, no maleficencia, autonomía, justicia y explicabilidad,

subrayando que la efectividad de estos principios depende de su incorporación en procesos organizacionales concretos, y no de meras declaraciones de intención.

La legitimidad organizacional, tal como la aborda Suchman (1995), se refiere a la percepción generalizada de que las acciones de una organización son deseables, apropiadas o adecuadas dentro de un sistema de normas, valores y creencias socialmente construido. Aplicada al contexto de la inteligencia artificial, esta noción sugiere que la adopción de sistemas algorítmicos sin un alineamiento normativo y ético adecuado tiende a generar desconfianza institucional y resistencia entre los grupos de interés afectados, socavando la legitimidad pragmática, moral y cognitiva de la organización. En sectores fuertemente regulados como las finanzas, la salud y la energía, la gobernanza responsable de los sistemas de inteligencia artificial funciona como un mecanismo de protección reputacional y de cumplimiento de marcos regulatorios cada vez más exigentes, entre ellos el Reglamento General de Protección de Datos de la Unión Europea y, en Brasil, la Ley General de Protección de Datos.

Desempeño ESG y creación de valor multidimensional

El desempeño ESG ha dejado de tratarse, en la literatura de gestión estratégica, como una métrica periférica asociada a las relaciones públicas o a la filantropía corporativa, y ha pasado a integrarse en las discusiones centrales sobre gestión de riesgos y creación de valor a largo plazo (Eccles, Ioannou & Serafeim, 2014; Serafeim, 2014). Eccles, Ioannou y Serafeim (2014), en un estudio comparativo entre empresas de alto y bajo desempeño en sostenibilidad, encuentran evidencia de que la adopción de políticas orientadas a la sostenibilidad corporativa se asocia con un desempeño de mercado y contable superior a largo plazo, especialmente cuando estas políticas pasan a formar parte de la cultura organizacional en lugar de permanecer como programas aislados de responsabilidad social corporativa. Serafeim (2014) argumenta además que la integración de la sostenibilidad en los procesos organizacionales afecta la calidad de la toma de decisiones al ampliar el conjunto de información considerada relevante para la gestión de riesgos y oportunidades.

La creación sostenible de valor, en este sentido, no proviene de simplemente añadir prácticas sociales y ambientales a modelos de negocio preexistentes, sino de conciliar el desempeño económico con la generación de valor para múltiples grupos de interés, incluidos trabajadores, comunidades, proveedores y el medio ambiente (Freeman, 1984; Porter & Kramer, 2006).

Porter y Kramer (2006) introducen el concepto de valor compartido para describir estrategias que generan simultáneamente valor económico para la empresa y valor social para la comunidad en la que opera, argumentando que esta integración representa una fuente significativa de innovación y diferenciación competitiva.

Incorporar la inteligencia artificial a este debate exige cautela conceptual. La tecnología es, ante todo, una herramienta de optimización y procesamiento de información. Sus efectos sobre el desempeño ESG y la creación de valor dependen del propósito estratégico, los valores institucionales y las directrices normativas establecidas por el liderazgo organizacional. Sin una orientación explícita hacia la sostenibilidad, los sistemas de inteligencia artificial tienden a optimizar variables de corto plazo, como la reducción de costos y las ganancias de productividad, en detrimento de resultados ambientales y sociales más difíciles de medir (Amit & Zott, 2001; Dwivedi et al., 2021). Este hallazgo refuerza la necesidad de un modelo teórico que precise las condiciones organizacionales bajo las cuales la inteligencia artificial contribuye genuinamente al desempeño ESG, en lugar de suponer esta contribución como un resultado automático de la adopción tecnológica.

ENFOQUE METODOLÓGICO

Este estudio se caracteriza como un ensayo teórico de naturaleza integradora (Torraco, 2016). Una revisión integradora se diferencia de las revisiones sistemáticas tradicionales en que permite la síntesis conceptual de literaturas provenientes de campos disciplinarios distintos, con el objetivo de proponer nuevas perspectivas teóricas en lugar de simplemente sintetizar hallazgos empíricos ya establecidos. Torraco (2016) sostiene que este tipo de revisión resulta particularmente adecuado para temas emergentes, en los que la literatura aún se encuentra fragmentada entre distintas tradiciones de investigación, situación que describe con razonable precisión el estado actual de la investigación sobre inteligencia artificial, ESG y creación de valor.

El desarrollo del modelo teórico siguió cuatro pasos. En el primero, se realizó un mapeo de los supuestos presentes en estudios que relacionan directamente la adopción de tecnologías digitales con resultados de sostenibilidad corporativa, identificando vacíos relacionados con la ausencia de mediadores organizacionales en dichas formulaciones. En el segundo paso, se articularon los constructos centrales de la teoría de las capacidades dinámicas, la literatura sobre gobernanza de la inteligencia artificial y los

estudios sobre desempeño ESG, buscando puntos de convergencia conceptual entre estos campos. En el tercer paso, se identificaron y conceptualizaron las capacidades organizacionales que, según la literatura consultada, funcionan como eslabones entre la herramienta tecnológica y los resultados sociales y ambientales buscados por las organizaciones. En el cuarto paso, se consolidó el modelo proposicional, vinculando las capacidades mediadoras identificadas con los mecanismos de riesgo y destrucción de valor discutidos en la quinta sección.

Como todo ensayo teórico, este artículo presenta limitaciones metodológicas importantes. La ausencia de datos empíricos impide la verificación estadística de las relaciones propuestas, que deben entenderse como proposiciones teóricas abiertas a futuras investigaciones y no como resultados comprobados. Además, aunque la literatura consultada buscó abarcar contribuciones relevantes de distintas tradiciones teóricas, su selección refleja las elecciones interpretativas del autor, que pueden no agotar la diversidad de perspectivas existentes sobre el tema. Estas limitaciones se retoman en la séptima sección, cuando se propone una agenda de investigación orientada a la validación empírica del modelo.

Vale destacar una advertencia metodológica adicional. Los ensayos teóricos corren el riesgo de presentar sus proposiciones como si ya estuvieran respaldadas por evidencia empírica consolidada, cuando en realidad representan hipótesis de trabajo derivadas del razonamiento deductivo y de analogías con literaturas afines. A lo largo de este artículo se procuró mantener clara la distinción entre afirmaciones respaldadas por investigaciones empíricas específicas y debidamente referenciadas, e inferencias teóricas propuestas por el autor a partir de la articulación de distintas vertientes conceptuales. Esta distinción es especialmente relevante en un campo aún en formación, en el que términos como gobernanza de la inteligencia artificial y capacidades organizacionales para la IA carecen de definiciones operacionales ampliamente aceptadas, lo que dificulta las comparaciones directas entre estudios y refuerza la necesidad de futuros esfuerzos de estandarización conceptual y metodológica.

EL MODELO TEÓRICO DE LAS CAPACIDADES ORGANIZACIONALES MEDIADORAS

La premisa central de este modelo es que el impacto de la inteligencia artificial sobre el desempeño ESG y la creación sostenible de valor no ocurre de manera directa, sino que está mediado por cuatro capacidades organizacionales interdependientes: la integración y calidad de los datos, el

aprendizaje organizacional, la gobernanza responsable y la conversión de información en decisiones estratégicas sostenibles. La debilidad en cualquiera de estas dimensiones compromete la cadena de valor asociada al uso de la tecnología, pudiendo convertir inversiones sustanciales en fuentes de pasivos operativos, legales y reputacionales.

Esta proposición se apoya en un argumento relativamente sencillo, pero con implicaciones significativas para la estrategia corporativa: los activos tecnológicos se han vuelto, en gran medida, imitables. El software de análisis predictivo, las plataformas de procesamiento de lenguaje natural y los servicios de aprendizaje automático están disponibles, en condiciones contractuales similares, para empresas de distintos tamaños y sectores. La diferenciación estratégica no reside, por lo tanto, en el software adquirido, sino en las rutinas organizacionales desarrolladas para gobernar, interpretar y aplicar críticamente los resultados generados por los sistemas algorítmicos. Las siguientes subsecciones detallan cada una de las cuatro capacidades propuestas.

Integración y calidad de los datos

La primera capacidad mediadora se refiere a la capacidad organizacional de recopilar, procesar, estructurar e integrar bases de datos heterogéneas, que abarcan tanto la información financiera y operativa tradicional como las métricas sociales y ambientales. Los sistemas de inteligencia artificial dependen, de manera esencial, de datos sólidos, precisos y representativos; la máxima bien establecida en la ciencia de datos según la cual datos de baja calidad producen resultados de baja calidad se aplica directamente a este contexto. La ausencia de una curaduría rigurosa compromete los modelos predictivos y distorsiona los diagnósticos estratégicos, con consecuencias que se extienden a la propia medición del desempeño ESG.

Las organizaciones que mantienen los datos dispersos en silos departamentales enfrentan dificultades sistemáticas para monitorear las emisiones indirectas de gases de efecto invernadero asociadas a la cadena de valor (alcance 3), auditar a los proveedores en busca de prácticas análogas al trabajo forzoso, o medir los indicadores internos de diversidad e inclusión. Sin integración de datos, los modelos predictivos se vuelven poco confiables o producen diagnósticos distorsionados de la sostenibilidad de la organización. Desarrollar esta capacidad requiere una inversión continua en arquitectura de datos, gobernanza de metadatos y cumplimiento de los marcos de protección de datos, incluidas la Ley General de Protección de

Datos de Brasil y el Reglamento General de Protección de Datos de la Unión Europea.

Cabe destacar un aspecto adicional: integrar los datos financieros y socioambientales en una misma plataforma analítica permite identificar correlaciones entre ecoeficiencia y rentabilidad que, de otro modo, permanecerían ocultas en análisis compartimentados. Esta integración ayuda a superar la dicotomía, todavía presente en parte de la práctica gerencial, entre el desempeño financiero y la responsabilidad social y ambiental, al tratarlas como dimensiones complementarias de una única estrategia de creación de valor. La calidad de los datos funciona así como el fundamento sobre el que se apoyan las demás capacidades mediadoras; sin esta base, las capacidades subsiguientes operan sobre un sustento informacional frágil.

Aprendizaje organizacional y capacidad de absorción

La tecnología de inteligencia artificial evoluciona en ciclos de innovación acelerados, lo que exige a las organizaciones desarrollar continuamente su capacidad de absorción, entendida por Cohen y Levinthal (1990) como la capacidad de reconocer el valor de la información externa nueva, asimilarla y aplicarla con fines comerciales. Esta capacidad no se limita a la adquisición de conocimiento técnico especializado; implica también cultivar habilidades analíticas y pensamiento crítico entre empleados, especialistas sectoriales y gestores de distintos niveles jerárquicos.

El aprendizaje organizacional garantiza que la empresa no permanezca como una usuaria pasiva de soluciones comerciales, sino que se vuelva capaz de interpretar, con profundidad analítica, los resultados generados por los modelos predictivos, comprender sus márgenes de error, identificar limitaciones metodológicas y ajustar los procesos operativos y estratégicos con base en las lecciones extraídas del funcionamiento de los sistemas. Cohen y Levinthal (1990) señalan que la capacidad de absorción depende del conocimiento previo relacionado, lo que sugiere que las organizaciones con una trayectoria de inversión en formación técnica y analítica tienden a encontrar más fácil absorber y aplicar correctamente las nuevas tecnologías de inteligencia artificial. Este hallazgo tiene implicaciones significativas para las empresas más pequeñas o con estructuras de capacitación menos desarrolladas, que pueden enfrentar dificultades adicionales para convertir la adopción tecnológica en un aprendizaje organizacional efectivo.

El aprendizaje organizacional cumple también una función de retroalimentación respecto a las demás capacidades del modelo. Las

auditorías de sesgos revelan necesidades específicas de capacitación; las fallas operativas identificadas en el funcionamiento de los sistemas motivan revisiones de las políticas de gobernanza; y la participación de distintas áreas funcionales en la evaluación de los sistemas amplía la capacidad de la organización para identificar riesgos que los equipos técnicos, actuando solos, podrían no percibir. Este proceso acerca la gestión de los sistemas de inteligencia artificial a un proceso continuo de aprendizaje institucional, en el que métricas, decisiones y consecuencias se comparan y revisan periódicamente.

Gobernanza responsable de la inteligencia artificial

La tercera capacidad mediadora corresponde a la capacidad organizacional de establecer directrices normativas de cumplimiento, explicabilidad técnica, mitigación del sesgo algorítmico y responsabilización institucional. Esta capacidad garantiza que los modelos utilizados por la organización respeten los derechos fundamentales, eviten reproducir discriminación sistémica en procesos sensibles como la selección de currículos, el otorgamiento de crédito o la asignación de recursos, y mantengan registros de auditoría accesibles para los reguladores y el público externo.

La gobernanza responsable se distingue de las declaraciones formales de principios éticos por su dimensión operacional. No basta con que una organización publique directrices generales sobre el uso ético de la inteligencia artificial; estas directrices deben traducirse en procesos concretos de evaluación de riesgos, documentación de los sistemas, pruebas de equidad y mecanismos efectivos para que las personas afectadas por decisiones automatizadas puedan impugnarlas. La ausencia de esta traducción operacional explica, en gran medida, por qué organizaciones con códigos de ética robustos siguen registrando incidentes graves relacionados con sesgos algorítmicos y fallas de responsabilización (Jobin et al., 2019).

Conversión de información en decisiones estratégicas sostenibles

La cuarta capacidad mediadora representa la etapa final y, en cierto sentido, más decisiva del modelo: la capacidad gerencial de traducir los análisis algorítmicos en decisiones tácticas y estratégicas orientadas a la creación de valor compartido. No basta con disponer de paneles sofisticados para la previsión de riesgos climáticos, proyecciones de eficiencia energética o modelos de valoración de riesgo social. La organización necesita agilidad decisoria, capital político interno y flexibilidad estructural para reasignar recursos financieros, rediseñar portafolios de productos bajo criterios de

economía circular y ajustar sus relaciones con los grupos de interés relevantes.

Esta capacidad es, con frecuencia, la más descuidada en las inversiones corporativas en inteligencia artificial, que tienden a concentrar recursos en la adquisición de tecnología y la formación de equipos técnicos, sin una inversión equivalente en procesos decisorios capaces de absorber y aplicar los resultados generados por esa tecnología. El resultado observado en muchos casos es la producción de análisis técnicamente robustos que permanecen desconectados de la estrategia organizacional, sin efecto práctico sobre las decisiones de inversión, la asignación de recursos o las relaciones con los grupos de interés.

Orquestación de capacidades y complementariedades

Las cuatro capacidades descritas no producen resultados de manera aislada. La calidad de los datos puede aumentar la precisión de los modelos analíticos, pero no garantiza que sus recomendaciones se conviertan en decisiones coherentes con los compromisos ambientales, sociales y de gobernanza. Del mismo modo, una estructura formal de gobernanza algorítmica puede convertirse en una mera cuestión de protocolo cuando la organización carece de profesionales capaces de interpretar las limitaciones estadísticas, cuestionar los supuestos y traducir la evidencia en decisiones estratégicas. El modelo asume, por lo tanto, la complementariedad entre las cuatro dimensiones: el valor surge de la combinación de infraestructura informacional, aprendizaje, control institucional y capacidad decisoria, y no de la presencia aislada de cualquiera de estas dimensiones.

Esta complementariedad ayuda a explicar por qué organizaciones que utilizan tecnologías similares, a menudo adquiridas a los mismos proveedores, obtienen resultados distintos en términos de desempeño ESG. Los recursos tecnológicos pueden comprarse en el mercado; las rutinas que permiten integrarlos a la estrategia, a los sistemas de incentivos y a los mecanismos de responsabilización dependen del aprendizaje acumulado a lo largo del tiempo, lo que confiere a estas rutinas un carácter tácito y de difícil imitación, en consonancia con los argumentos clásicos de la teoría de las capacidades dinámicas (Teece, 2007).

La interdependencia entre las capacidades también produce efectos de retroalimentación relevantes para la gobernanza organizacional. Las decisiones sostenibles bien documentadas mejoran la calidad de los datos disponibles para futuras rondas de análisis; los incidentes identificados en las auditorías revelan necesidades específicas de capacitación; y la

participación de distintas áreas funcionales, como jurídico, sostenibilidad y operaciones, amplía la capacidad de identificar riesgos que los equipos técnicos, por sí solos, podrían no percibir. La organización deja de evaluar únicamente la precisión técnica de sus modelos y pasa también a examinar la calidad social, ambiental y económica de las decisiones que esos modelos respaldan.

Madurez organizacional y trayectoria de desarrollo

Las capacidades mediadoras pueden entenderse como dimensiones de madurez organizacional que evolucionan con el tiempo en respuesta a la experiencia acumulada, los cambios regulatorios y las decisiones de inversión. En las etapas iniciales, la organización tiende a concentrarse en la adquisición de herramientas y en la ejecución de proyectos piloto, con una integración limitada entre áreas y una fuerte dependencia de proveedores externos. Los datos permanecen dispersos, la evaluación de riesgos ocurre de forma reactiva, y las decisiones se orientan principalmente por indicadores operativos de corto plazo. En esta etapa, la inteligencia artificial puede generar ganancias de eficiencia localizadas, pero es poco probable que modifique de manera consistente el desempeño ESG de la organización o su proceso de creación de valor.

En una etapa intermedia, la organización comienza a formalizar políticas, asignar responsabilidades específicas y ampliar la alfabetización analítica de sus gestores. Los proyectos empiezan a evaluarse según criterios de relevancia estratégica, y la empresa desarrolla mecanismos para documentar modelos, monitorear resultados y revisar periódicamente las decisiones. Una cooperación más estrecha entre las áreas de tecnología, sostenibilidad, jurídico y negocios reduce la fragmentación característica de la etapa anterior, aunque todavía pueden persistir iniciativas paralelas, métricas incompatibles entre áreas y disputas sobre la propiedad de los datos. El principal desafío en esta etapa consiste en convertir las estructuras formales de gobernanza en rutinas genuinamente incorporadas al día a día de la organización.

La madurez avanzada se caracteriza por la incorporación de la gobernanza de la inteligencia artificial al sistema de gestión regular de la organización. El liderazgo sénior monitorea sistemáticamente los riesgos y oportunidades, los modelos se supervisan a lo largo de todo su ciclo de vida, y los resultados se evalúan de manera integrada en las dimensiones económica, ambiental y social. La organización desarrolla la capacidad de comparar alternativas tecnológicas, discontinuar sistemas inadecuados cuando es necesario, y

aprender de los incidentes que ocurren. En esta etapa, la inteligencia artificial deja de tratarse como una iniciativa excepcional y pasa a formar parte de la arquitectura decisoria de la empresa, sujeta a los mismos principios de responsabilización, control y creación de valor aplicados a otros activos estratégicos.

Esta evolución, sin embargo, no sigue necesariamente una trayectoria lineal. Los cambios regulatorios, la rotación de liderazgo, los procesos de fusión y adquisición o los incidentes graves pueden acelerar o interrumpir el desarrollo de estas capacidades. Además, una misma organización puede presentar alta madurez en ciertas áreas y fragilidad marcada en otras. Las instituciones financieras, por ejemplo, tienden a desarrollar controles robustos para los modelos de otorgamiento de crédito, un área sujeta a regulación específica y escrutinio continuo, mientras utilizan al mismo tiempo sistemas poco supervisados en áreas como recursos humanos o marketing, donde la presión regulatoria históricamente ha sido menor. Evaluar la madurez organizacional debe, por lo tanto, considerar procesos y aplicaciones específicos, evitando clasificaciones genéricas basadas únicamente en la existencia de políticas corporativas declaradas.

La trayectoria de desarrollo de estas capacidades también debe ser compatible con la materialidad de los riesgos involucrados en cada aplicación. No todas las organizaciones necesitan construir estructuras de gobernanza idénticas, pero es fundamental que la intensidad de los controles corresponda al impacto potencial de las decisiones automatizadas. Las empresas más pequeñas pueden adoptar mecanismos de gobernanza simplificados, siempre que preserven una documentación mínima, una responsabilización clara y una posibilidad concreta de revisar las decisiones tomadas con el apoyo de sistemas algorítmicos. El objetivo no es reproducir estructuras burocráticas complejas en todos los contextos, sino garantizar que el ritmo de la innovación tecnológica no supere la capacidad institucional de comprender y gestionar sus consecuencias.

Relaciones con proveedores y dependencia tecnológica

La mayoría de las organizaciones adopta la inteligencia artificial a través de plataformas, servicios en la nube y modelos desarrollados por terceros, en lugar de un desarrollo interno completo. Esta estructura amplía el acceso a tecnologías avanzadas, incluso para empresas más pequeñas, pero crea relaciones de dependencia que afectan directamente la gobernanza, la seguridad de la información y la capacidad de aprendizaje organizacional. Cuando los contratos de suministro no garantizan documentación adecuada,

portabilidad de datos ni información transparente sobre las actualizaciones del sistema, la organización contratante pierde la autonomía para evaluar críticamente la herramienta que utiliza y corregir los problemas identificados. La responsabilidad por los impactos generados, sin embargo, permanece en la organización que utiliza la tecnología en sus procesos internos, independientemente de quién la haya desarrollado.

La selección de proveedores tecnológicos debe incluir, de manera más sistemática que en la práctica corporativa actual, criterios técnicos y ESG combinados. Más allá del desempeño técnico, el costo y la disponibilidad, es necesario evaluar las prácticas de protección de datos, el consumo de recursos computacionales y energéticos, la gestión de sesgos en los modelos ofrecidos, la transparencia respecto a las limitaciones conocidas y las condiciones laborales a lo largo de la cadena de suministro tecnológico del proveedor. Los contratos deben establecer obligaciones claras de notificación de incidentes, cambios relevantes en el sistema y limitaciones técnicas conocidas, además de garantizar el acceso a evidencia suficiente con fines de auditoría, respetando los derechos de propiedad intelectual del proveedor.

La dependencia tecnológica excesiva puede reducirse mediante estrategias como la arquitectura modular, la adopción de estándares interoperables, la capacitación interna continua y el mantenimiento de un nivel mínimo de conocimiento técnico sobre el funcionamiento de los procesos automatizados utilizados por la organización. No es necesario que la empresa desarrolle internamente todos los componentes tecnológicos que utiliza, pero es recomendable que conserve la competencia suficiente para comprender el funcionamiento general de las soluciones adoptadas, interpretar críticamente sus resultados y sustituir proveedores cuando sea necesario sin comprometer la continuidad de sus operaciones. Esta capacidad protege la continuidad operativa de la organización y reduce el riesgo de que las decisiones estratégicas queden excesivamente condicionadas por los intereses comerciales de los proveedores externos.

MECANISMOS DE DESTRUCCIÓN DE VALOR Y RIESGOS DE LA IA EN CONTEXTOS ESG

La ausencia, atrofia o debilitamiento de las capacidades organizacionales mediadoras propuestas en este modelo puede convertir la adopción de la inteligencia artificial en un vector de destrucción de valor corporativo y de erosión del desempeño ESG. La literatura consultada permite identificar cinco mecanismos principales asociados a este riesgo: la opacidad decisorio,

los sesgos y la discriminación sistémica, el consumo intensivo de recursos computacionales, la baja responsabilización y la desalineación estratégica. Las siguientes subsecciones examinan cada uno de estos mecanismos con mayor profundidad.

Opacidad, explicabilidad y riesgo decisorio

La opacidad algorítmica no es simplemente una cuestión de la dificultad técnica para comprender modelos complejos, especialmente aquellos basados en redes neuronales profundas. También resulta de decisiones organizacionales relacionadas con la documentación, el acceso a la información y la distribución de responsabilidades dentro de la empresa. Un sistema puede ser técnicamente explicable, en el sentido de que sus técnicas permiten interpretar cómo se produjo una determinada decisión, y aun así permanecer institucionalmente opaco, cuando sus criterios no se comunican a los grupos de interés afectados o cuando ningún órgano organizacional es responsable de revisar sus resultados. La explicabilidad que importa para la gobernanza corporativa debe, por lo tanto, combinar comprensión técnica, justificación gerencial y una posibilidad concreta de impugnación por parte de las personas afectadas por las decisiones (Floridi et al., 2018).

Para las decisiones de alto impacto, es recomendable que la organización registre el propósito del sistema utilizado, los datos empleados en su entrenamiento, las métricas de desempeño adoptadas, las limitaciones conocidas del modelo y las condiciones que requieren intervención humana directa. Este tipo de registro protege a la organización frente a la falsa impresión de neutralidad algorítmica, muchas veces utilizada para justificar decisiones que, en la práctica, reflejan elecciones técnicas e institucionales específicas. El registro también permite reconstruir el proceso decisorio durante auditorías internas y externas, algo especialmente relevante cuando las decisiones automatizadas afectan el acceso al crédito, las oportunidades de empleo, el acceso a servicios esenciales o la exposición a riesgos ambientales.

Sesgos, representatividad e impactos sociales

Los sesgos observados en los sistemas de inteligencia artificial no resultan únicamente de errores de programación, aunque estos también pueden ocurrir. Con frecuencia reflejan desigualdades históricas presentes en los datos utilizados para entrenar los modelos, elecciones inadecuadas de variables, muestras que representan deficientemente a ciertos grupos poblacionales, u objetivos de optimización que ignoran consecuencias distributivas relevantes (Buolamwini & Gebu, 2018). Una organización

puede mejorar la eficiencia promedio de un determinado proceso al mismo tiempo que empeora sus resultados para grupos específicos, situación que pasa inadvertida cuando la evaluación del desempeño considera únicamente promedios agregados.

Por esta razón, la evaluación del desempeño de los sistemas de inteligencia artificial debe desagregarse siempre que exista un riesgo significativo de trato desigual, permitiendo comparar las tasas de error, los beneficios y los costos entre los distintos grupos afectados. La gobernanza responsable exige que la organización defina de antemano qué diferencias de trato son aceptables, cuáles requieren corrección inmediata y cuáles hacen que un determinado sistema sea incompatible con sus compromisos ESG declarados. Este juicio involucra valores, derechos y prioridades estratégicas, y por lo tanto no debe delegarse exclusivamente en el equipo técnico responsable del desarrollo de los modelos. La participación de profesionales de cumplimiento, sostenibilidad, recursos humanos y jurídico reduce la probabilidad de que decisiones técnicamente eficientes produzcan efectos socialmente injustos, convirtiendo el control de sesgos en una responsabilidad organizacional compartida y no en una tarea aislada del área de tecnología.

Huella ambiental y eficiencia computacional

El potencial ambiental positivo de la inteligencia artificial suele asociarse, en la literatura aplicada y en la práctica corporativa, con la optimización energética, el mantenimiento predictivo de equipos y el monitoreo de emisiones. Estos beneficios, sin embargo, deben sopesarse sistemáticamente frente a los recursos consumidos por el propio sistema. Los modelos de gran escala requieren una capacidad computacional significativa, electricidad, agua para la refrigeración de los centros de datos y una renovación frecuente de equipos, con impactos ambientales que, en muchos casos, permanecen en gran medida invisibles para la organización que utiliza el sistema (Strubell, Ganesh & McCallum, 2019). La evaluación ambiental de los sistemas de inteligencia artificial no debe considerar únicamente la etapa de entrenamiento del modelo, sino todo el ciclo de vida, incluyendo el almacenamiento de datos, la ejecución de inferencias, las actualizaciones periódicas y el descarte del hardware.

La capacidad de integración de datos discutida anteriormente puede contribuir a este tipo de control al vincular los indicadores de desempeño técnico con los datos de consumo energético y los resultados operativos. En lugar de adoptar automáticamente el modelo más complejo disponible en el

mercado, la organización debe seleccionar la solución proporcional al problema que busca resolver, evaluando si las ganancias marginales de precisión justifican los costos ambientales y financieros adicionales asociados a los modelos de mayor escala. Delmas y Burbano (2011) advierten, en un contexto más amplio, sobre el riesgo de que iniciativas presentadas como sostenibles produzcan, en la práctica, impactos ocultos incompatibles con las metas climáticas corporativas, fenómeno conocido como lavado verde (greenwashing). Este riesgo se aplica directamente a los proyectos de inteligencia artificial promovidos bajo argumentos de sostenibilidad sin una evaluación rigurosa de sus costos ambientales reales.

Responsabilización y supervisión humana

Delegar tareas decisorias a sistemas automatizados no elimina la responsabilidad humana por los resultados de esas decisiones. Por el contrario, exige que la organización defina con mayor precisión quién autoriza el uso de un determinado sistema, quién supervisa su desempeño a lo largo del tiempo, quién tiene la autoridad para desactivarlo y quién es formalmente responsable de cualquier daño causado por su funcionamiento. Expresiones genéricas como «la decisión del algoritmo», todavía comunes en las justificaciones organizacionales, tienden a diluir la responsabilidad y debilitar la gobernanza, sugiriendo una autonomía técnica que, en la práctica, no existe de forma aislada de los procesos organizacionales que definieron e implementaron el sistema. La tecnología debe permanecer incorporada dentro de una cadena decisoria identificable, con autoridades claras, límites y procedimientos de escalamiento en caso de fallas.

La supervisión humana efectiva no consiste simplemente en incluir formalmente a una persona en la etapa final del proceso decisorio. Cuando el profesional responsable carece de información suficiente, tiempo adecuado o autonomía real para anular la recomendación generada automáticamente, su participación se vuelve meramente nominal, sin capacidad real de corrección. La supervisión debe diseñarse de manera que la persona responsable comprenda el fundamento de la recomendación presentada, reconozca situaciones excepcionales que requieran un análisis adicional y pueda suspender el proceso decisorio sin enfrentar incentivos organizacionales que desalienten esa decisión. Esta condición depende de la inversión en capacitación, del diseño adecuado de roles y responsabilidades, y de una cultura organizacional que fomente el cuestionamiento de las recomendaciones automatizadas, en lugar de tratarlas como decisiones finales e incuestionables.

Desalineación y destrucción estratégica de valor

Los proyectos de inteligencia artificial pueden destruir valor incluso cuando funcionan correctamente desde el punto de vista técnico. Esto ocurre, por ejemplo, cuando el sistema aborda problemas de baja relevancia estratégica, automatiza procesos que deberían rediseñarse en lugar de simplemente acelerarse, o desvía recursos financieros y gerenciales de iniciativas organizacionales más importantes. El entusiasmo tecnológico, a menudo presente en las discusiones corporativas sobre inteligencia artificial, puede llevar a una organización a definir la solución tecnológica antes de comprender adecuadamente la necesidad que pretende atender, dando lugar a inversiones difíciles de integrar y a resultados escasamente relacionados con la estrategia general de la empresa. El análisis de materialidad ESG, entendido como la identificación sistemática de los temas relevantes para un determinado sector y para los grupos de interés de la organización, ofrece un filtro útil para priorizar aplicaciones de inteligencia artificial capaces de abordar riesgos y oportunidades genuinamente significativos, en lugar de proyectos impulsados principalmente por modas de mercado.

El alineamiento estratégico entre los proyectos de inteligencia artificial y la agenda ESG requiere indicadores que vayan más allá de las medidas convencionales de productividad y reducción de costos. Un proyecto de gestión de la cadena de suministro, por ejemplo, debería evaluarse también por su contribución real a la trazabilidad de los productos, la mejora de las condiciones laborales a lo largo de la cadena y la reducción de los impactos ambientales asociados a la producción y la logística. Un sistema utilizado en la gestión de personas debe considerar, además de los indicadores de eficiencia, la calidad de las decisiones que produce, sus efectos sobre la diversidad de la fuerza laboral, la confianza de los empleados y el bienestar organizacional. La creación sostenible de valor depende de incorporar estas dimensiones desde el diseño inicial del proyecto de inteligencia artificial, en lugar de añadirlas posteriormente como un elemento complementario en los informes de sostenibilidad producidos después de que el sistema ya ha sido implementado.

IMPLICACIONES GERENCIALES Y ESTRATÉGICAS

Las reflexiones desarrolladas a lo largo de este artículo ofrecen implicaciones relevantes para ejecutivos, miembros de consejos de administración y gestores de sostenibilidad. En primer lugar, ponen de relieve los límites de los enfoques tecnocéntricos que tratan la inteligencia

artificial como una solución ya lista, desvinculada del contexto institucional en el que será utilizada. Invertir en tecnología sin una inversión equivalente en el fortalecimiento de las capacidades organizacionales mediadoras discutidas en este artículo equivale, en la práctica, a acelerar procesos que pueden ser ineficientes o insostenibles, solo que más rápido y con menos margen para una corrección oportuna. En segundo lugar, el modelo sugiere que gestionar la sostenibilidad en la era digital requiere la creación de instancias interdisciplinarias de gobernanza algorítmica, que reúnan a especialistas en datos, profesionales de cumplimiento ESG y líderes estratégicos capaces de participar en la curaduría continua de los datos utilizados, las auditorías periódicas de sesgos y la validación ética de las decisiones respaldadas por sistemas automatizados.

Arquitectura de gobernanza y responsabilidades

Institucionalizar las capacidades propuestas en este modelo requiere una arquitectura de gobernanza compatible con el tamaño, el sector y el nivel de riesgo de cada organización. El consejo de administración debe establecer principios generales y supervisar los riesgos significativos asociados al uso de la inteligencia artificial, mientras que la alta dirección traduce estas directrices en políticas concretas, asignación de recursos y definición de prioridades. Los comités interdisciplinarios pueden contribuir de manera significativa a la evaluación de proyectos específicos, pero no deben sustituir la responsabilidad de las áreas de negocio propietarias de los procesos en los que se utilizan los sistemas. Todo sistema relevante necesita un responsable de negocio identificable, un responsable técnico y una función de control independiente capaz de cuestionar, cuando sea necesario, las decisiones tomadas por los dos primeros.

Una política corporativa de inteligencia artificial debe establecer criterios claros de clasificación de riesgos, documentación mínima requerida, procedimientos para la adquisición de soluciones externas, protección de datos, pruebas previas a la implementación, supervisión humana continua y criterios objetivos para la discontinuación de sistemas. Los sistemas de bajo impacto pueden seguir procedimientos de gobernanza simplificados, mientras que las aplicaciones que afectan derechos fundamentales, la seguridad física o decisiones financieras significativas requieren una evaluación más rigurosa. Esta proporcionalidad evita sobrecargar de burocracia a las iniciativas de bajo riesgo, sin reducir el control necesario sobre los usos potencialmente dañinos de la tecnología.

Ciclo de vida de los sistemas de inteligencia artificial

La gobernanza responsable debe acompañar todo el ciclo de vida del sistema, no solo su fase inicial de implementación. En la fase de diseño, deben definirse claramente el problema que el sistema pretende resolver, su propósito específico, los grupos de interés potencialmente afectados y los criterios objetivos de éxito que se utilizarán en su evaluación. Durante el desarrollo, es necesario realizar pruebas sistemáticas de calidad, robustez técnica, seguridad de la información y equidad entre los distintos grupos afectados. Antes de la puesta en producción, la organización debe evaluar los impactos potenciales y preparar procedimientos de contingencia para escenarios de falla. Después de que el sistema entra en operación, su desempeño debe monitorearse de manera continua para identificar cambios en la distribución de los datos, deterioro progresivo de la precisión o efectos no anticipados en fases anteriores.

La discontinuación de sistemas es también una parte significativa de este ciclo de vida. Los sistemas que han perdido su propósito original, que dependen de tecnologías obsoletas o que presentan riesgos desproporcionados frente a los beneficios generados no deben permanecer en operación simplemente por la inversión ya realizada en su implementación, un razonamiento que constituye una forma reconocida del sesgo de costos hundidos. La capacidad organizacional de desactivar o sustituir una solución tecnológica demuestra madurez estratégica y reduce la tendencia a perpetuar decisiones pasadas independientemente de su adecuación actual. Los registros detallados de versiones, cambios realizados e incidentes ocurridos permiten trasladar las lecciones aprendidas en un determinado proyecto a las iniciativas futuras de la organización.

Indicadores para evaluar el valor sostenible

La medición de los resultados obtenidos con los sistemas de inteligencia artificial debe combinar indicadores técnicos, económicos y ESG, evitando un enfoque restringido a métricas puramente operativas. Los indicadores de precisión, tiempo de procesamiento y disponibilidad del sistema son necesarios, pero insuficientes para evaluar el valor real del sistema para la organización y sus grupos de interés. Es recomendable que la organización también monitoree el consumo de recursos computacionales y energéticos, la distribución de errores entre los distintos grupos afectados, las quejas recibidas, las decisiones revertidas tras impugnación, la confianza percibida de empleados y clientes, y los efectos observados sobre grupos de interés relevantes específicos. Este conjunto ampliado de indicadores permite identificar situaciones en las que ganancias operativas aparentes ocultan un

deterioro social o ambiental no captado por las métricas técnicas convencionales.

El valor efectivamente creado por un determinado sistema debe evaluarse en relación con una línea base clara y con soluciones alternativas plausibles para el mismo problema. No toda mejora observada resulta necesariamente de la inteligencia artificial adoptada, y no toda tarea organizacional requiere realmente automatización para llevarse a cabo con calidad adecuada. Las comparaciones sistemáticas con procesos anteriores, los proyectos piloto bien documentados y las evaluaciones periódicas reducen el riesgo de atribuir erróneamente resultados a la tecnología cuando otros factores organizacionales podrían explicar las mejoras observadas. La calidad de esta medición también depende del reconocimiento de efectos de largo plazo, como la dependencia excesiva de proveedores tecnológicos específicos, la pérdida de conocimiento técnico interno resultante de la externalización de funciones críticas, o cambios significativos en las relaciones de la organización con trabajadores y consumidores.

PROPOSICIONES TEÓRICAS Y AGENDA DE INVESTIGACIÓN

El modelo desarrollado a lo largo de este artículo permite formular proposiciones teóricas capaces de orientar futuros estudios empíricos. La primera proposición establece que la relación entre la adopción de la inteligencia artificial y el desempeño ESG tiende a ser positiva únicamente cuando la organización cuenta con niveles adecuados de integración y calidad en los datos utilizados por sus sistemas. La segunda proposición sugiere que el aprendizaje organizacional amplía la capacidad de la empresa para traducir los resultados analíticos en cambios concretos en los procesos y en las decisiones estratégicas. La tercera proposición sostiene que la gobernanza responsable reduce la incidencia de los mecanismos de destrucción de valor discutidos en la quinta sección, asociados a la opacidad, los sesgos y la baja responsabilización institucional.

La cuarta proposición establece que la capacidad de convertir información en decisiones sostenibles media la relación entre la capacidad analítica organizacional y la creación de valor. Una empresa puede producir previsiones técnicamente robustas sin que ello se traduzca en cambios en sus inversiones o prácticas de gestión; en este caso, la inteligencia generada por el sistema permanece desconectada de la estrategia organizacional a pesar de su calidad técnica. La quinta proposición indica que las cuatro capacidades ejercen un efecto complementario entre sí: una deficiencia marcada en cualquiera de ellas tiende a limitar los beneficios producidos por

las demás, incluso cuando estas últimas están bien desarrolladas. Esta hipótesis de complementariedad puede examinarse en futuras investigaciones mediante modelos configuracionales o análisis comparativos cualitativos entre organizaciones de distintos sectores y tamaños.

Futuras investigaciones podrían desarrollar escalas específicas para medir cada una de las cuatro capacidades propuestas y someter el modelo a prueba empírica en distintos contextos sectoriales. Las instituciones financieras, las industrias intensivas en recursos naturales, las empresas tecnológicas y las organizaciones de salud presentan perfiles de riesgo y estructuras regulatorias suficientemente distintos como para justificar investigaciones específicas en cada uno de estos contextos. Los estudios longitudinales permitirían observar cómo evolucionan las capacidades organizacionales a lo largo del tiempo en respuesta a incidentes significativos, cambios regulatorios o procesos de sustitución tecnológica. Los estudios de caso en profundidad podrían, a su vez, detallar los procesos internos de aprendizaje organizacional y los conflictos frecuentemente observados entre las áreas técnicas, las áreas de sostenibilidad y la alta dirección.

Una vía adicional de investigación consiste en indagar los resultados distributivos asociados a la adopción de la inteligencia artificial en contextos ESG. Es posible que la creación de valor se concentre predominantemente en los accionistas, mientras que costos significativos se trasladan a trabajadores, consumidores o comunidades afectadas por las operaciones de la organización. La evaluación empírica de este fenómeno debe identificar explícitamente quién recibe los beneficios generados por la tecnología, quién soporta los riesgos asociados a su uso y cómo los mecanismos de gobernanza discutidos en este artículo alteran esta distribución a lo largo del tiempo. Este enfoque amplía el análisis más allá del desempeño agregado de la organización y acerca la investigación sobre inteligencia artificial y ESG a debates más amplios sobre legitimidad organizacional, justicia distributiva y responsabilidad corporativa, temas centrales para la gestión estratégica contemporánea.

CONSIDERACIONES FINALES

La relación entre la inteligencia artificial, el desempeño ESG y la creación de valor corporativo constituye uno de los temas más relevantes para la teoría y la práctica de la gestión estratégica contemporánea. Como este artículo ha buscado demostrar, el despliegue de activos tecnológicos avanzados no garantiza, por sí solo, la sostenibilidad corporativa ni una

ventaja competitiva duradera. El valor económico y socioambiental generado por la aplicación de la inteligencia artificial en contextos ESG está condicionado de manera bastante directa por la solidez de cuatro capacidades organizacionales mediadoras: la integración rigurosa de los datos, el aprendizaje institucional continuo, la gobernanza algorítmica responsable y la capacidad gerencial de convertir los análisis técnicos en decisiones estratégicas sostenibles.

Cuando estas competencias están ausentes o poco desarrolladas en una organización, la tecnología tiende a actuar como un amplificador de riesgos preexistentes, generando opacidad decisoria, sesgos discriminatorios y destrucción de valor en lugar de creación de valor. Este hallazgo tiene implicaciones prácticas significativas para los gestores: sugiere que las inversiones en inteligencia artificial deben evaluarse no solo por su potencial técnico, sino también por la capacidad organizacional disponible para absorberlas de manera responsable. La distinción entre adquirir tecnología y desarrollar una competencia genuina para utilizarla de manera sostenible, central en el argumento presentado en este artículo, resulta particularmente relevante en un momento en que muchas organizaciones enfrentan presiones para adoptar soluciones de inteligencia artificial a un ritmo acelerado, muchas veces sin tiempo adecuado para desarrollar las capacidades necesarias para hacerlo de manera responsable.

El modelo propuesto no pretende ofrecer respuestas definitivas sobre cómo cada organización debe estructurar su gobernanza de la inteligencia artificial, ni busca establecer un conjunto único de prácticas aplicables indiscriminadamente a distintos sectores y contextos. Su propósito es más modesto: ofrecer un marco conceptual capaz de orientar tanto la reflexión académica como la práctica gerencial respecto a las condiciones organizacionales necesarias para que la inteligencia artificial contribuya genuinamente al desempeño ESG y a la creación sostenible de valor, en lugar de suponer esta contribución como un resultado automático de la adopción tecnológica.

Como agenda para futuras investigaciones, se recomienda la validación empírica de este modelo mediante enfoques metodológicos diversos, incluyendo el modelado de ecuaciones estructurales para poner a prueba las relaciones propuestas entre las capacidades mediadoras y los resultados organizacionales, así como estudios de caso comparativos entre distintos sectores industriales, capaces de profundizar en la comprensión de los matices sectoriales de la gobernanza algorítmica orientada a la sostenibilidad. Se espera que este ensayo teórico contribuya a desplazar el

debate sobre inteligencia artificial y sostenibilidad corporativa de una perspectiva centrada exclusivamente en la tecnología hacia una centrada en las capacidades organizacionales necesarias para utilizarla de manera responsable, distinción que resulta cada vez más relevante a medida que estas tecnologías se convierten en una parte estructural de la estrategia corporativa contemporánea.

REFERENCIAS

Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal*, 22(6-7), 493-520. <https://doi.org/10.1002/smj.187>

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99-120. <https://doi.org/10.1177/014920639101700108>

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

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128-152. <https://doi.org/10.2307/2393553>

Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, 54(1), 64-87. <https://doi.org/10.1525/cmr.2011.54.1.64>

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>

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., & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal*, 21(10-11), 1105-1121. [https://doi.org/10.1002/1097-0266\(200010/11\)21:10/11<1105::AID-SMJ133>3.0.CO;2-E](https://doi.org/10.1002/1097-0266(200010/11)21:10/11<1105::AID-SMJ133>3.0.CO;2-E)

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.

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>

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, 28(9), 1435-1448. <https://doi.org/10.1177/0170840607081138>

Porter, M. E., & Kramer, M. R. (2006). Strategy and society: The link between competitive advantage and corporate social responsibility. *Harvard Business Review*, 84(12), 78-92.

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>

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645-3650. <https://doi.org/10.18653/v1/P19-1355>

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571-610. <https://doi.org/10.2307/258788>

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>

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509-533. [https://doi.org/10.1002/\(SICI\)1097-0266\(199708\)18:7<509::AID-SMJ882>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z)

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>

Wamba, S. F., Bawack, R. E., Guthrie, C., Queiroz, M. M., & Carillo, K. D. A. (2021). Are we preparing for a good AI society? A bibliometric review and research agenda. *Technological Forecasting and Social Change*, 164, 120482. <https://doi.org/10.1016/j.techfore.2020.120482>

Whittaker, M. (2021). The steep cost of capture. *Interactions*, 28(6), 50-55. <https://doi.org/10.1145/3488666>

CONTRIBUCIÓN DEL AUTOR

Lincon Lopes: conceptualización, metodología, investigación, análisis formal, redacción del manuscrito original, revisión y edición, y supervisión. La curaduría de datos y la obtención de financiamiento no aplican a este estudio teórico-integrador, que no utilizó un conjunto de datos primario ni recibió financiamiento externo.

NOTA BIOGRÁFICA

APROBACIÓN ÉTICA

No aplica. El estudio es de naturaleza teórico-integradora y no involucró participantes humanos, datos personales identificables ni experimentación con animales.

DISPONIBILIDAD DE LOS DATOS

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

DECLARACIONES EDITORIALES

Financiamiento: el estudio no recibió financiamiento externo.

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

Uso de inteligencia artificial: no declarado por el autor.

© 2026 El autor. Publicado por Noemica bajo la Licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0).