Noemica Journal

noemicajournal.org

e-ISSN: 3086-7037

ARTICLE

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

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

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

**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

| SUBMITTED | ACCEPTED | PUBLISHED | DOI |
|---|---|---|---|
| 02/04/2024 | 18/10/2024 | 06/12/2024 | in process |

## 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.

**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.

## INTRODUCTION

The corporate adoption of technologies based on artificial intelligence has been reshaping, at an accelerating pace, the foundations of competitive strategy, organizational architecture, and business models on a global scale. At the same time, institutional, regulatory, and market pressures associated with the ESG (Environmental, Social, and Governance) agenda require companies to demonstrate genuine commitment to environmental sustainability, social equity, and governance integrity, beyond the declaratory rhetoric found in institutional reports. This twofold movement, marked by the intense digitalization of operations and by growing demands for measurable social and environmental outcomes, has led managers and researchers to assume, often uncritically, that incorporating advanced algorithmic tools produces, almost automatically, gains in ESG performance and corporate value creation.

The available evidence does not simply support this assumption. Acquiring artificial intelligence systems, hiring specialized consultancies, or implementing isolated machine learning models does not, on its own, ensure consistent improvements in an organization's environmental, social, and governance indicators. On the contrary, a considerable portion of the critical literature on technology and society documents significant externalities associated with these systems: high energy consumption in data centers (Strubell et al., 2019), the reproduction of discriminatory biases in automated decisions (Buolamwini & Gebru, 2018), opacity in decision-making processes, and erosion of organizational legitimacy among diverse stakeholder groups (Suchman, 1995; Whittaker, 2021). The relationship between the adoption of artificial intelligence and sustainable value creation is, therefore, neither direct nor automatic. It is mediated by managerial routines, intangible assets, and specific organizational capabilities that determine whether the technology will be converted into legitimate competitive advantage or into an additional source of risk.

The question guiding this theoretical essay can be formulated as follows: which organizational capabilities mediate the relationship between the adoption of artificial intelligence and ESG performance, conditioning the net creation or destruction of corporate value? It is argued throughout the text that converting the technical potential of artificial intelligence into sustainable outcomes depends on the deliberate development of internal competencies aimed at data integration and quality, continuous organizational learning, responsible algorithmic governance, and the capacity to translate technical analyses into strategic decisions oriented simultaneously toward economic value and social and environmental value.

The purpose of this article is to propose a theoretical model that explains how these organizational capabilities condition the relationship between artificial intelligence, ESG performance, and value creation. To this end, the text is structured into eight sections, in addition to this introduction. The second section presents the theoretical framework, articulating dynamic capabilities theory, the literature on artificial intelligence governance, and studies on ESG performance. The third section describes the methodological approach adopted, which is theoretical-integrative in nature. The fourth section sets out the proposed model and its four mediating capabilities, also discussing their orchestration, the organizational maturity required to sustain them, and the risks associated with dependence on technology vendors. The fifth section examines the mechanisms of value destruction linked to the use of artificial intelligence in ESG contexts. The sixth section discusses managerial and strategic implications, including governance architecture and management of the systems' life cycle. The seventh section presents theoretical propositions and a research agenda. Finally, the eighth section offers concluding remarks.

An initial observation is warranted. This article does not aim to exhaust the discussion on artificial intelligence in organizations, nor does it propose a definitive model. Its purpose is more modest and, at the same time, more specific: to offer a conceptual framework that helps researchers and managers distinguish between the instrumental use of technology and the building of organizational competencies genuinely capable of sustaining consistent environmental, social, and governance outcomes over time.

## THEORETICAL FRAMEWORK

The theoretical framework of this study articulates three bodies of literature that, although often treated separately, maintain close conceptual relationships when applied to the problem of artificial intelligence adoption in ESG contexts: dynamic capabilities theory, studies on artificial intelligence governance and algorithmic ethics, and the literature on ESG performance and value creation. This articulation stems from a simple observation: models that treat technology as an exogenous and neutral productivity factor tend to underestimate the role of organizational choices in determining its effects. Artificial intelligence systems are sociotechnical artifacts whose outcomes depend both on their technical properties and on the routines, values, and decision-making structures of the organizations that implement them (Orlikowski, 2007). This premise guides the construction of the model presented in the fourth section.

### *Dynamic capabilities and competitive advantage in the digital economy*

The Resource-Based View holds that a firm's competitive advantage stems from possessing resources that are valuable, rare, difficult to imitate, and organizationally exploited (Barney, 1991). This perspective, however, presents significant limitations in environments of rapid technological change, in which the mere possession of resources does not guarantee sustained superior performance. Dynamic capabilities theory, developed by Teece, Pisano, and Shuen (1997) and later refined by Teece (2007), addresses this limitation by shifting the focus from the static possession of resources to the organizational ability to integrate, build, and reconfigure them in response to changes in the competitive environment. According to Teece (2007), this ability unfolds into three microfoundations: sensing capacity, aimed at identifying opportunities and threats; seizing capacity, related to mobilizing resources to capture the value identified; and transforming capacity, associated with the continuous reconfiguration of assets, structures, and organizational routines.

Eisenhardt and Martin (2000) contribute to this discussion by pointing out that dynamic capabilities do not, in themselves, constitute a direct source of sustainable competitive advantage; rather, they function as organizational processes that allow the firm to reconfigure its resource base consistently with changes in the environment. This distinction is relevant to the argument of this article. Advanced technological resources, including artificial intelligence systems, originate essentially as operational efficiency tools. Their transformation into a lasting competitive differentiator depends on the development of higher-order organizational routines capable of absorbing the technology, adapting it to specific decision-making processes, and realigning it with the demands of regulators, investors, and other stakeholders.

Applying this logic to the context of artificial intelligence, it can be argued that the technology alone does not constitute a source of sustainable competitive advantage, nor of superior ESG performance. Machine learning systems, analytical platforms, and automation tools are now widely available on the market, often as cloud-based services offered by a small number of large vendors. This availability reduces the differentiation potential associated with the mere acquisition of the technology (Wamba et al., 2021). The competitive differentiator thus shifts to the organizational routines that allow a firm to use these resources consistently with its strategy and with the expectations of its stakeholders, an argument underlying the proposition of the mediating capabilities discussed in the fourth section of this article.

It is worth noting that this shift in focus, from the possession of resources to the ability to orchestrate them, has practical consequences for how boards of directors and investors assess corporate artificial intelligence projects.

Announcements of technology investment tend to be received in capital markets as a positive signal, associated with expectations of efficiency and innovation gains. Dynamic capabilities theory suggests caution regarding this reading: the value of the investment depends less on the amount allocated to hardware, software, or model licensing, and more on the prior existence, or parallel development, of organizational routines capable of absorbing this technology and reconfiguring processes around it. Investors and board members who evaluate only the volume of resources allocated to artificial intelligence, without examining the underlying organizational maturity, run the risk of systematically overestimating the value-creation potential of these investments.

### *Artificial intelligence governance, algorithmic ethics, and organizational legitimacy*

Traditional corporate governance was largely conceived to address agency problems between shareholders and executives, with an emphasis on financial and accounting control mechanisms. The growing presence of algorithmic systems in corporate decision-making processes broadens the scope of this governance, requiring mechanisms capable of overseeing automated decisions that are often opaque and difficult to audit. The literature on artificial intelligence ethics and governance has become established as a relevant field for strategic management, bringing together contributions on ethical principles applicable to autonomous systems, technical explainability, and accountability mechanisms (Floridi et al., 2018; Jobin, Ienca & Vayena, 2019).

Jobin, Ienca, and Vayena (2019), in a review of guidance documents on artificial intelligence ethics published by governments, companies, and international organizations, identify convergence around principles such as transparency, fairness, non-maleficence, accountability, and privacy, while also pointing to significant divergences regarding how these principles should be operationalized. This finding is important for the argument of this article: the existence of declared principles is not equivalent to the existence of organizational capabilities capable of applying them consistently. Floridi et al. (2018) reinforce this point by proposing an ethical framework for artificial intelligence systems that articulates beneficence, non-maleficence, autonomy, justice, and explicability, stressing that the effectiveness of these principles depends on their incorporation into concrete organizational processes, rather than mere statements of intent.

Organizational legitimacy, as discussed by Suchman (1995), refers to the generalized perception that an organization's actions are desirable, proper, or appropriate within a socially constructed system of norms, values, and beliefs. Applied to the context of artificial intelligence, this notion suggests that adopting algorithmic systems without adequate normative and ethical alignment tends to generate institutional distrust and resistance among affected stakeholders, undermining the organization's pragmatic, moral, and cognitive legitimacy. In heavily regulated sectors such as finance, healthcare, and energy, the responsible governance of artificial intelligence systems functions as a mechanism for reputational protection and compliance with increasingly stringent regulatory frameworks, including the European Union's General Data Protection Regulation and, in Brazil, the General Data Protection Law.

### *ESG performance and multidimensional value creation*

ESG performance has ceased to be treated, in the strategic management literature, as a peripheral metric associated with public relations or corporate philanthropy, and has instead become integrated into central discussions on risk management and long-term value creation (Eccles, Ioannou & Serafeim, 2014; Serafeim, 2014). Eccles, Ioannou, and Serafeim (2014), in a comparative study of high- and low-sustainability-performing firms, find evidence that adopting policies aimed at corporate sustainability is associated with superior market and accounting performance over the long run, especially when these policies become part of the organizational culture rather than remaining isolated corporate social responsibility programs. Serafeim (2014) further argues that integrating sustainability into organizational processes affects the quality of decision-making by expanding the set of information considered relevant for managing risks and opportunities.

Sustainable value creation, in this sense, does not stem from simply adding social and environmental practices to pre-existing business models, but from reconciling economic performance with value generation for multiple stakeholders, including workers, communities, suppliers, and the environment (Freeman, 1984; Porter & Kramer, 2006). Porter and Kramer (2006) introduce the concept of shared value to describe strategies that simultaneously generate economic value for the firm and social value for the community in which it operates, arguing that this integration represents a significant source of innovation and competitive differentiation.

Bringing artificial intelligence into this debate requires conceptual caution. The technology is, primarily, a tool for optimization and information processing. Its effects on ESG performance and value creation depend on the strategic purpose, institutional values, and normative guidelines set by organizational leadership. Without explicit direction toward sustainability, artificial intelligence systems tend to optimize short-term variables, such as cost reduction and productivity gains, at the expense of environmental and social outcomes that are harder to measure (Amit & Zott, 2001; Dwivedi et al., 2021). This finding reinforces the need for a theoretical model that spells out the organizational conditions under which artificial intelligence genuinely contributes to ESG performance, rather than assuming this contribution as an automatic result of technology adoption.

## METHODOLOGICAL APPROACH

This study is characterized as a theoretical essay of an integrative nature (Torraco, 2016). An integrative review differs from traditional systematic reviews in that it allows for the conceptual synthesis of literatures drawn from distinct disciplinary fields, with the aim of proposing new theoretical perspectives rather than merely synthesizing previously established empirical findings. Torraco (2016) argues that this type of review is particularly suited to emerging topics, in which the literature is still fragmented across different research traditions, a situation that reasonably accurately describes the current state of research on artificial intelligence, ESG, and value creation.

The development of the theoretical model followed four steps. In the first, a mapping was carried out of the assumptions present in studies that directly relate the adoption of digital technologies to corporate sustainability outcomes, identifying gaps related to the absence of organizational mediators in these formulations. In the second step, the central constructs of dynamic capabilities theory, the literature on artificial intelligence governance, and studies on ESG performance were articulated, seeking points of conceptual convergence among these fields. In the third step, the organizational capabilities that, according to the literature consulted, function as links between the technological tool and the social and environmental outcomes sought by organizations were identified and conceptualized. In the fourth step, the propositional model was consolidated, linking the mediating capabilities identified to the mechanisms of risk and value destruction discussed in the fifth section.

noemicajournal.org
e-ISSN: 3086-7037

As with any theoretical essay, this article presents significant methodological limitations. The absence of empirical data prevents statistical verification of the proposed relationships, which should be understood as theoretical propositions open to future investigation rather than as proven results. In addition, although the literature consulted sought to encompass relevant contributions from different theoretical traditions, its selection reflects the author's interpretive choices, which may not exhaust the diversity of existing perspectives on the topic. These limitations are revisited in the seventh section, when a research agenda aimed at the empirical validation of the model is proposed.

An additional methodological caveat is worth noting. Theoretical essays run the risk of presenting their propositions as though already backed by consolidated empirical evidence, when in fact they represent working hypotheses derived from deductive reasoning and analogies with related literatures. Throughout this article, an effort was made to keep clear the distinction between claims supported by specific, duly referenced empirical research and theoretical inferences proposed by the author from the articulation of different conceptual strands. This distinction is especially relevant in a field still taking shape, in which terms such as artificial intelligence governance and organizational capabilities for AI lack widely accepted operational definitions, which hinders direct comparisons between studies and reinforces the need for future efforts toward conceptual and methodological standardization.

## THE THEORETICAL MODEL OF MEDIATING ORGANIZATIONAL CAPABILITIES

The central premise of this model is that the impact of artificial intelligence on ESG performance and sustainable value creation does not occur directly, but is mediated by four interdependent organizational capabilities: data integration and quality, organizational learning, responsible governance, and the conversion of information into sustainable strategic decisions. Weakness in any one of these dimensions compromises the value chain associated with the use of the technology, potentially turning substantial investments into sources of operational, legal, and reputational liabilities.

This proposition rests on a relatively simple argument with significant implications for corporate strategy: technological assets have largely become imitable. Predictive analytics software, natural language processing platforms, and machine learning services are available, under similar contractual terms, to firms of different sizes and sectors. Strategic differentiation, therefore, does not lie in the software acquired, but in the organizational routines developed to govern, interpret, and critically apply the results generated by algorithmic systems. The following subsections detail each of the four proposed capabilities.

#### *Data integration and quality*

The first mediating capability refers to the organizational ability to collect, process, structure, and integrate heterogeneous databases, including both traditional financial and operational information and social and environmental metrics. Artificial intelligence systems depend, in essential ways, on data that are sound, accurate, and representative; the well-established maxim in data science that low-quality data produces low-quality results applies directly to this context. The absence of rigorous curation undermines predictive models and distorts strategic diagnoses, with consequences that extend to the very measurement of ESG performance.

Organizations that keep data scattered across departmental silos face systematic difficulties in monitoring indirect greenhouse gas emissions associated with the value chain (scope 3), auditing suppliers for practices analogous to forced labor, or measuring internal diversity and inclusion indicators. Without data integration, predictive models become unreliable or produce distorted diagnoses of the organization's sustainability. Developing this capability requires ongoing investment in data architecture, metadata governance, and compliance with data protection frameworks, including the Brazilian General Data Protection Law and the European Union's General Data Protection Regulation.

There is an additional aspect worth highlighting: integrating financial and social/environmental data on the same analytical platform makes it possible to identify correlations between eco-efficiency and profitability that would otherwise remain hidden in siloed analyses. This integration helps overcome the dichotomy, still present in part of managerial practice, between financial performance and social and environmental responsibility, treating them instead as complementary dimensions of a single value-creation strategy. Data quality thus functions as the foundation on which the other mediating capabilities rest; without this foundation, the subsequent capabilities operate on a fragile informational base.

#### *Organizational learning and absorptive capacity*

Artificial intelligence technology evolves in rapid innovation cycles, which requires organizations to continuously develop absorptive capacity, understood by Cohen and Levinthal (1990) as the ability to recognize the value of new, external information, assimilate it, and apply it to commercial ends. This capacity is not limited to acquiring specialized technical knowledge; it also involves cultivating analytical skills and critical thinking among employees, sector specialists, and managers at different hierarchical levels.

Organizational learning ensures that the firm does not remain a passive user of commercial solutions, but instead becomes capable of interpreting, with analytical depth, the results generated by predictive models, understanding their margins of error, identifying methodological limitations, and adjusting operational and strategic processes based on the lessons drawn from the systems' operation. Cohen and Levinthal (1990) note that absorptive capacity depends on related prior knowledge, which suggests that organizations with a track record of investment in technical and analytical training tend to find it easier to absorb and correctly apply new artificial intelligence technologies. This finding has significant implications for smaller firms or those with less developed training structures, which may face additional difficulties in converting technology adoption into effective organizational learning.

Organizational learning also serves a feedback function in relation to the other capabilities in the model. Bias audits reveal specific training needs; operational failures identified in system operation prompt reviews of governance policies; and the involvement of different functional areas in system evaluation broadens the organization's ability to identify risks that technical teams, working alone, might not perceive. This process brings the management of artificial intelligence systems closer to a continuous process of institutional learning, in which metrics, decisions, and consequences are periodically compared and reviewed.

### *Responsible governance of artificial intelligence*

The third mediating capability corresponds to the organizational ability to establish normative guidelines for compliance, technical explainability, mitigation of algorithmic bias, and institutional accountability. This capability ensures that the models used by the organization respect fundamental rights, avoid reproducing systemic discrimination in sensitive processes such as résumé screening, credit granting, or resource allocation, and maintain audit trails accessible to regulators and the external public.

Responsible governance is distinguished from formal declarations of ethical principles by its operational dimension. It is not enough for an organization to publish general guidelines on the ethical use of artificial intelligence; these guidelines must translate into concrete processes of risk assessment, system documentation, fairness testing, and effective mechanisms for individuals affected by automated decisions to contest them. The absence of this operational translation largely explains why organizations with robust codes of ethics still record serious incidents related to algorithmic bias and accountability failures (Jobin et al., 2019).

#### *Converting information into sustainable strategic decisions*

The fourth mediating capability represents the final and, in a sense, most decisive stage of the model: the managerial ability to translate algorithmic analyses into tactical and strategic decisions oriented toward shared value creation. Having sophisticated dashboards for climate risk forecasting, energy efficiency projections, or social risk pricing models is not enough. The organization needs decision-making agility, internal political capital, and structural flexibility to reallocate financial resources, redesign product portfolios under circular economy criteria, and adjust its relationships with relevant stakeholders.

This capability is, frequently, the most neglected in corporate investments in artificial intelligence, which tend to concentrate resources on acquiring technology and building technical teams, without an equivalent investment in decision-making processes capable of absorbing and applying the results generated by that technology. The result observed in many cases is the production of technically robust analyses that remain disconnected from organizational strategy, with no practical effect on investment decisions, resource allocation, or stakeholder relationships.

#### *Orchestration of capabilities and complementarities*

The four capabilities described do not produce results in isolation. Data quality may increase the precision of analytical models, but it does not ensure that their recommendations are converted into decisions consistent with environmental, social, and governance commitments. Likewise, a formal structure of algorithmic governance can become merely a matter of protocol when the organization lacks professionals capable of interpreting statistical limitations, questioning assumptions, and translating evidence into strategic choices. The model therefore assumes complementarity among the four dimensions: value emerges from the combination of informational infrastructure, learning, institutional control, and decision-making capacity, and not from the isolated presence of any one of these dimensions.

This complementarity helps explain why organizations that use similar technologies, often purchased from the same vendors, achieve different results in terms of ESG performance. Technological resources can be bought on the market; the routines that allow them to be integrated into strategy, incentive systems, and accountability mechanisms depend on learning accumulated over time, which gives these routines a tacit character that is difficult to imitate, consistent with the classic arguments of dynamic capabilities theory (Teece, 2007).

The interdependence among capabilities also produces feedback effects relevant to organizational governance. Well-documented sustainable decisions improve the quality of data available for future rounds of analysis; incidents identified in audits reveal specific training needs; and the involvement of different functional areas, such as legal, sustainability, and operations, broadens the ability to identify risks that technical teams might not perceive on their own. The organization ceases to evaluate only the technical accuracy of its models and begins to also examine the social, environmental, and economic quality of the decisions those models support.

### *Organizational maturity and development trajectory*

The mediating capabilities can be understood as dimensions of organizational maturity that evolve over time in response to accumulated experience, regulatory changes, and investment decisions. In early stages, the organization tends to focus on acquiring tools and running pilot projects, with limited integration across areas and strong dependence on external vendors. Data remain scattered, risk assessment occurs reactively, and decisions are driven mainly by short-term operational indicators. At this stage, artificial intelligence may generate localized efficiency gains, but it is unlikely to consistently change the organization's ESG performance or its value-creation process.

At an intermediate stage, the organization begins to formalize policies, assign specific responsibilities, and expand the analytical literacy of its managers. Projects begin to be evaluated according to strategic relevance criteria, and the firm develops mechanisms to document models, track results, and periodically review decisions. Closer cooperation among technology, sustainability, legal, and business areas reduces the fragmentation characteristic of the previous stage, although parallel initiatives, incompatible metrics across areas, and disputes over data ownership may still persist. The main challenge at this stage is turning formal governance structures into routines that are genuinely embedded in day-to-day organizational life.

Advanced maturity is characterized by the incorporation of artificial intelligence governance into the organization's regular management system. Senior leadership systematically monitors risks and opportunities, models are monitored throughout their entire life cycle, and results are evaluated in an integrated way across economic, environmental, and social dimensions. The organization develops the capacity to compare technological alternatives, discontinue inadequate systems when necessary, and learn from incidents that occur. At this stage, artificial intelligence ceases to be treated as an exceptional initiative and becomes part of the firm's decision-making architecture, subject to the same principles of accountability, control, and value creation applied to other strategic assets.

This evolution, however, does not necessarily follow a linear trajectory. Regulatory changes, leadership turnover, merger and acquisition processes, or serious incidents can accelerate or interrupt the development of these capabilities. Moreover, the same organization may display high maturity in certain areas and marked fragility in others. Financial institutions, for example, tend to develop robust controls for credit-granting models, an area subject to specific regulation and continuous scrutiny, while at the same time using poorly supervised systems in areas such as human resources or marketing, where regulatory pressure has historically been lower. Assessing organizational maturity should therefore consider specific processes and applications, avoiding generic classifications based solely on the existence of declared corporate policies.

The development trajectory of these capabilities also needs to be compatible with the materiality of the risks involved in each application. Not every organization needs to build identical governance structures, but it is essential that the intensity of controls match the potential impact of automated decisions. Smaller firms may adopt simplified governance mechanisms, provided they preserve minimal documentation, clear accountability, and a concrete possibility of reviewing decisions made with the support of algorithmic systems. The goal is not to reproduce complex bureaucratic structures in every context, but to ensure that the pace of technological innovation does not outstrip the institutional capacity to understand and manage its consequences.

***Vendor relationships and technological dependence***

Most organizations adopt artificial intelligence through platforms, cloud services, and models developed by third parties, rather than through full in-house development. This structure broadens access to advanced technologies, including for smaller firms, but creates relationships of dependence that directly affect governance, information security, and organizational learning capacity. When supply contracts do not guarantee adequate documentation, data portability, or transparent information about system updates, the contracting organization loses the autonomy to critically evaluate the tool it uses and to correct identified problems. Responsibility for the impacts generated, however, remains with the organization that uses the technology in its internal processes, regardless of who developed it.

The selection of technology vendors should include, more systematically than currently occurs in corporate practice, combined technical and ESG criteria. Beyond technical performance, cost, and availability, it is necessary to assess data protection practices, computational and energy resource consumption, bias management in the models offered, transparency regarding known limitations, and working conditions along the vendor's technology supply chain. Contracts need to establish clear obligations for reporting incidents, relevant system changes, and known technical limitations, as well as ensuring access to sufficient evidence for audit purposes, while respecting the vendor's intellectual property rights.

Excessive technological dependence can be reduced through strategies such as modular architecture, adoption of interoperable standards, ongoing internal training, and maintaining a minimum level of technical knowledge about how the automated processes used by the organization work. It is not necessary for the firm to develop in-house every technological component it uses, but it is advisable that it retain sufficient competence to understand the general workings of the solutions adopted, critically interpret their results, and replace vendors when necessary without compromising the continuity of its operations. This capability protects the organization's operational continuity and reduces the risk that strategic decisions become excessively conditioned by the commercial interests of external vendors.

## MECHANISMS OF VALUE DESTRUCTION AND AI RISKS IN ESG CONTEXTS

The absence, atrophy, or weakening of the mediating organizational capabilities proposed in this model can turn the adoption of artificial intelligence into a vector of corporate value destruction and ESG performance erosion. The literature consulted allows for the identification of five main mechanisms associated with this risk: decision opacity, bias and systemic discrimination, intensive computational resource consumption, weak accountability, and strategic misalignment. The following subsections examine each of these mechanisms in greater depth.

### *Opacity, explainability, and decision-making risk*

Algorithmic opacity is not simply a matter of the technical difficulty of understanding complex models, especially those based on deep neural networks. It also results from organizational decisions related to documentation, access to information, and the distribution of responsibilities within the firm. A system can be technically explainable, in the sense that its techniques allow one to interpret how a given decision was produced, and still remain institutionally opaque, when its criteria are not communicated to affected stakeholders or when no organizational body is responsible for reviewing its results. Explainability that matters for corporate governance must therefore combine technical understanding, managerial justification, and a concrete possibility of contestation by individuals affected by the decisions (Floridi et al., 2018).

For high-impact decisions, it is advisable that the organization record the purpose of the system used, the data employed in its training, the performance metrics adopted, the model's known limitations, and the conditions requiring direct human intervention. This type of record protects the organization against the false impression of algorithmic neutrality, often used to justify decisions that, in practice, reflect specific technical and institutional choices. The record also makes it possible to reconstruct the decision-making process during internal and external audits, something especially relevant when automated decisions affect access to credit, employment opportunities, access to essential services, or exposure to environmental risks.

### *Bias, representativeness, and social impacts*

The biases observed in artificial intelligence systems do not result solely from programming errors, although such errors can also occur. They frequently reflect historical inequalities present in the data used to train the models, inadequate choices of variables, samples that poorly represent certain population groups, or optimization objectives that ignore relevant distributive consequences (Buolamwini & Gebru, 2018). An organization may improve the average efficiency of a given process while simultaneously worsening its outcomes for specific groups, a situation that goes unnoticed when performance evaluation considers only aggregate averages.

For this reason, the performance evaluation of artificial intelligence systems should be disaggregated whenever there is a significant risk of unequal treatment, allowing error rates, benefits, and costs to be compared across different affected groups. Responsible governance requires the organization to define in advance which differences in treatment are acceptable, which require immediate correction, and which make a given system incompatible with its declared ESG commitments. This judgment involves values, rights, and strategic priorities, and should therefore not be delegated exclusively to the technical team responsible for developing the models. The involvement of compliance, sustainability, human resources, and legal professionals reduces the likelihood that technically efficient decisions will produce socially unjust effects, making bias control a shared organizational responsibility rather than an isolated task of the technology area.

### *Environmental footprint and computational efficiency*

The positive environmental potential of artificial intelligence is often associated, in the applied literature and in corporate practice, with energy optimization, predictive equipment maintenance, and emissions monitoring. These benefits, however, need to be systematically weighed against the resources consumed by the system itself. Large-scale models require significant computational capacity, electricity, water for cooling data centers, and frequent equipment renewal, with environmental impacts that, in many cases, remain largely invisible to the organization using the system (Strubell, Ganesh & McCallum, 2019). The environmental assessment of artificial intelligence systems should not consider only the model training stage, but the entire life cycle, including data storage, inference execution, periodic updates, and hardware disposal.

The data integration capability discussed earlier can contribute to this type of control by linking technical performance indicators with energy consumption data and operational results. Rather than automatically adopting the most complex model available on the market, the organization should select the solution proportional to the problem it aims to solve, assessing whether marginal accuracy gains justify the additional environmental and financial costs associated with larger-scale models. Delmas and Burbano (2011) warn, in a broader context, of the risk that initiatives presented as sustainable may, in practice, produce hidden impacts incompatible with corporate climate targets, a phenomenon known as greenwashing. This risk applies directly to artificial intelligence projects promoted under sustainability arguments without a rigorous assessment of their real environmental costs.

### *Accountability and human oversight*

Delegating decision-making tasks to automated systems does not eliminate human responsibility for the outcomes of those decisions. On the contrary, it requires the organization to define more precisely who authorizes the use of a given system, who monitors its performance over time, who has the authority to shut it down, and who is formally accountable for any harm caused by its operation. Generic expressions such as "the algorithm's decision," still common in organizational justifications, tend to dilute responsibility and weaken governance, suggesting a technical autonomy that, in practice, does not exist in isolation from the organizational processes that defined and implemented the system. The technology must remain embedded within an identifiable decision-making chain, with clear authorities, limits, and escalation procedures in the event of failures.

Effective human oversight is not simply about formally including a person at the final step of the decision-making process. When the professional responsible lacks sufficient information, adequate time, or real autonomy to override the automatically generated recommendation, their involvement becomes merely nominal, without any real capacity for correction. Oversight needs to be designed so that the responsible person understands the basis of the recommendation presented, recognizes exceptional situations that require additional analysis, and can suspend the decision-making process without facing organizational incentives that discourage such a decision. This condition depends on investment in training, appropriate design of roles and responsibilities, and an organizational culture that encourages questioning automated recommendations, rather than treating them as final and unchallengeable decisions.

### *Misalignment and strategic value destruction*

Artificial intelligence projects can destroy value even when they function correctly from a technical standpoint. This happens, for example, when the system addresses problems of low strategic relevance, automates processes that should instead be redesigned rather than merely sped up, or diverts financial and managerial resources away from more important organizational initiatives. Technological enthusiasm, often present in corporate discussions about artificial intelligence, can lead an organization to define the technological solution before adequately understanding the need it is meant to address, resulting in investments that are difficult to integrate and outcomes only loosely related to the firm's overall strategy. ESG materiality analysis, understood as the systematic identification of topics relevant to a given sector and to the organization's stakeholders, offers a useful filter for prioritizing artificial intelligence applications capable of addressing genuinely significant risks and opportunities, rather than projects driven mainly by market fads.

Strategic alignment between artificial intelligence projects and the ESG agenda requires indicators that go beyond conventional measures of productivity and cost reduction. A supply chain management project, for example, should also be evaluated for its actual contribution to product traceability, improved working conditions along the chain, and reduced environmental impacts associated with production and logistics. A system used in people management needs to consider, in addition to efficiency indicators, the quality of the decisions it produces, its effects on workforce diversity, employee trust, and organizational well-being. Sustainable value creation depends on incorporating these dimensions into the initial design of the artificial intelligence project, rather than adding them later as a supplementary item in sustainability reports produced after the system has been implemented.

## MANAGERIAL AND STRATEGIC IMPLICATIONS

The reflections developed throughout this article offer relevant implications for executives, board members, and sustainability managers. First, they highlight the limits of technocentric approaches that treat artificial intelligence as a ready-made solution detached from the institutional context in which it will be used. Investing in technology without an equivalent investment in strengthening the mediating organizational capabilities discussed in this article amounts, in practice, to accelerating processes that may be inefficient or unsustainable, only faster and with less room for timely correction. Second, the model suggests that managing sustainability in the digital age requires the creation of interdisciplinary bodies for algorithmic governance, bringing together data specialists, ESG compliance professionals, and strategic leaders capable of engaging in the ongoing curation of the data used, periodic bias audits, and the ethical validation of decisions supported by automated systems.

### *Governance architecture and responsibilities*

Institutionalizing the capabilities proposed in this model requires a governance architecture compatible with each organization's size, sector, and risk level. The board of directors should establish general principles and oversee the significant risks associated with the use of artificial intelligence, while senior management translates these guidelines into concrete policies, resource allocation, and priority-setting. Interdisciplinary committees can contribute significantly to the evaluation of specific projects, but should not replace the responsibility of the business areas that own the processes in which the systems are used. Every relevant system needs an identifiable business owner, a technical owner, and an independent control function capable of challenging decisions made by the first two when necessary.

A corporate artificial intelligence policy should establish clear criteria for risk classification, minimum required documentation, procedures for acquiring external solutions, data protection, pre-implementation testing, ongoing human oversight, and objective criteria for discontinuing systems. Low-impact systems can follow simplified governance procedures, while applications that affect fundamental rights, physical safety, or significant financial decisions require more thorough evaluation. This proportionality avoids excessively burdening low-risk initiatives with bureaucracy, without reducing the necessary control over potentially harmful uses of the technology.

### *Life cycle of artificial intelligence systems*

Responsible governance needs to accompany the entire life cycle of the system, not just its initial implementation phase. In the design phase, the problem the system is intended to solve, its specific purpose, the potentially affected stakeholders, and the objective success criteria to be used in its evaluation should be clearly defined. During development, systematic testing of quality, technical robustness, information security, and fairness across different affected groups is necessary. Before deployment into production, the organization should assess potential impacts and prepare contingency procedures for failure scenarios. After the system enters operation, its performance needs to be continuously monitored in order to identify changes in data distribution, progressive deterioration in accuracy, or effects not anticipated in earlier phases.

The discontinuation of systems is also a significant part of this life cycle. Systems that have lost their original purpose, that rely on obsolete technologies, or that present risks disproportionate to the benefits generated should not remain in operation simply because of the investment already made in their implementation, a rationale that constitutes a recognized form of sunk-cost bias. The organizational capacity to shut down or replace a technological solution demonstrates strategic maturity and reduces the tendency to perpetuate past decisions regardless of their current adequacy. Detailed records of versions, changes made, and incidents that have occurred allow the lessons learned in a given project to be carried over to the organization's future initiatives.

### *Indicators for assessing sustainable value*

Measuring the results obtained with artificial intelligence systems should combine technical, economic, and ESG indicators, avoiding a narrow focus on purely operational metrics. Accuracy indicators, processing time, and system availability are necessary but insufficient for assessing the system's real value to the organization and its stakeholders. It is advisable for the organization to also track computational and energy resource consumption, the distribution of errors across different affected groups, complaints received, decisions reversed following contestation, the perceived trust of employees and customers, and observed effects on specific relevant stakeholders. This expanded set of indicators makes it possible to identify situations in which apparent operational gains conceal social or environmental deterioration not captured by conventional technical metrics.

The value actually created by a given system should be assessed relative to a clear baseline and to plausible alternative solutions to the same problem. Not every observed improvement necessarily results from the artificial intelligence adopted, and not every organizational task actually requires automation to be carried out with adequate quality. Systematic comparisons with prior processes, well-documented pilot projects, and periodic evaluations reduce the risk of wrongly attributing results to the technology when other organizational factors could explain the improvements observed. The quality of this measurement also depends on recognizing long-term effects, such as excessive dependence on specific technology vendors, loss of internal technical knowledge resulting from outsourcing critical functions, or significant changes in the organization's relationships with workers and consumers.

## THEORETICAL PROPOSITIONS AND RESEARCH AGENDA

The model developed throughout this article allows for the formulation of theoretical propositions capable of guiding future empirical studies. The first proposition establishes that the relationship between the adoption of artificial intelligence and ESG performance tends to be positive only when the organization has adequate levels of integration and quality in the data used by its systems. The second proposition suggests that organizational learning broadens the firm's capacity to translate analytical results into concrete changes in processes and strategic decisions. The third proposition holds that responsible governance reduces the incidence of the value-destruction mechanisms discussed in the fifth section, associated with opacity, bias, and weak institutional accountability.

The fourth proposition states that the capacity to convert information into sustainable decisions mediates the relationship between organizational analytical capacity and value creation. A firm may produce technically robust forecasts without, however, altering its investments or management practices as a result; in this case, the intelligence generated by the system remains disconnected from organizational strategy despite its technical quality. The fifth proposition indicates that the four capabilities have a complementary effect on one another: a marked deficiency in any one of them tends to limit the benefits produced by the others, even when those other capabilities are well developed. This complementarity hypothesis can be examined in future research through configurational models or qualitative comparative analyses among organizations of different sectors and sizes.

Future research could develop specific scales to measure each of the four proposed capabilities and empirically test the model across different sectoral contexts. Financial institutions, natural-resource-intensive industries, technology companies, and healthcare organizations present sufficiently distinct risk profiles and regulatory structures to warrant specific investigations in each of these contexts. Longitudinal studies would allow researchers to observe how organizational capabilities evolve over time in response to significant incidents, regulatory changes, or processes of technological replacement. In-depth case studies could, in turn, detail internal processes of organizational learning and the conflicts frequently observed among technical areas, sustainability areas, and senior management.

An additional avenue for research consists of investigating the distributive outcomes associated with the adoption of artificial intelligence in ESG contexts. It is possible that value creation accrues predominantly to shareholders, while significant costs are shifted to workers, consumers, or communities affected by the organization's operations. Empirical assessment of this phenomenon should explicitly identify who receives the benefits generated by the technology, who bears the risks associated with its use, and how the governance mechanisms discussed in this article alter this distribution over time. This approach broadens the analysis beyond the organization's aggregate performance and brings research on artificial intelligence and ESG closer to broader debates on organizational legitimacy, distributive justice, and corporate responsibility, themes that are central to contemporary strategic management.

## CONCLUDING REMARKS

The relationship among artificial intelligence, ESG performance, and corporate value creation constitutes one of the most relevant topics for the theory and practice of contemporary strategic management. As this article has sought to demonstrate, the deployment of advanced technological assets does not, on its own, guarantee corporate sustainability or lasting competitive advantage. The economic and social/environmental value generated by applying artificial intelligence in ESG contexts is quite directly conditioned by the robustness of four mediating organizational capabilities: rigorous data integration, continuous institutional learning, responsible algorithmic governance, and the managerial capacity to convert technical analyses into sustainable strategic decisions.

When these competencies are absent or underdeveloped in an organization, the technology tends to act as an amplifier of pre-existing risks, generating decision opacity, discriminatory biases, and value destruction rather than value creation. This finding has significant practical implications for managers: it suggests that investments in artificial intelligence should be evaluated not only for their technical potential, but also for the organizational capacity available to absorb them responsibly. The distinction between acquiring technology and developing genuine competence to use it sustainably, central to the argument presented in this article, appears particularly relevant at a time when many organizations face pressure to adopt artificial intelligence solutions at an accelerated pace, often without adequate time to develop the capabilities needed to do so responsibly.

The proposed model does not aim to offer definitive answers on how each organization should structure its artificial intelligence governance, nor does it seek to establish a single set of practices indiscriminately applicable across different sectors and contexts. Its purpose is more modest: to offer a conceptual framework capable of guiding both academic reflection and managerial practice regarding the organizational conditions necessary for artificial intelligence to genuinely contribute to ESG performance and sustainable value creation, rather than assuming this contribution to be an automatic result of technology adoption.

e-ISSN: 3086-7037

As an agenda for future research, empirical validation of this model is recommended through varied methodological approaches, including structural equation modeling to test the proposed relationships between the mediating capabilities and organizational outcomes, as well as comparative case studies across different industrial sectors, capable of deepening our understanding of sectoral nuances in algorithmic governance geared toward sustainability. It is hoped that this theoretical essay contributes to shifting the debate on artificial intelligence and corporate sustainability from a perspective centered exclusively on technology to one centered on the organizational capabilities needed to use it responsibly, a distinction that appears increasingly relevant as these technologies become a structural part of contemporary corporate strategy.

## REFERENCES

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

Floridi, L., Cowls, 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

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

## BIOGRAPHICAL NOTE

## AUTHOR CONTRIBUTION

Lincon Lopes: conceptualization, methodology, investigation, formal analysis, writing of the original manuscript, review and editing, and supervision. Data curation and funding acquisition do not apply to this theoretical-integrative study, which did not use a primary dataset and received no external funding.

#### EDITORIAL DECLARATIONS

Funding: the study received no external funding.

Conflict of interest: the author declares no conflict of interest.

Use of artificial intelligence: not declared by the author.

#### ETHICAL APPROVAL

Not applicable. The study is theoretical-integrative in nature and did not involve human participants, identifiable personal data, or animal experimentation.

#### DATA AVAILABILITY

Not applicable. The article did not produce a primary dataset.