

AI-Driven Sustainable Value Creation: Integrating Responsible Innovation and Stakeholder Perspectives

¹Dr. Pawan Kumar, ²Shivi, ³Gurinderjeet Kaur, ⁴Shardendu Bhushan Sinha, ⁵Anuradha Sinha, ⁶Rajni Vats, ⁷Ravinder Singh, ⁸Gourav Kamboj

¹Assistant professor of Law, Teerthanker Mahaveer University, Moradabad

²Assistant Professor in Commerce, CKD Institute of Management & Technology, Tarn Taran

³Assistant Professor, Department of Commerce and Management, CKD Institute of Management and Technology, Tarn Taran, Punjab

⁴Assistant Professor, Department of Management, I.T.S. Engineering College, Greater Noida

⁵Assistant Professor, Department of Management, I.T.S. Engineering College, Greater Noida

⁶Assistant Professor, Department of Management, Tilak Raj Chadha Institute of Management and Technology (TIMT), Yamunanagar

⁷Research scholar at Maharishi Markandeshwar Deemed to be University, Mullana, Ambala

⁸Assistant Professor at LM Thapar School of Management, Thapar University, Patiala, Punjab

Abstract

This paper investigates how artificial intelligence capabilities can be harnessed to create sustainable value for organizations and their stakeholders when guided by principles of responsible innovation and inclusive governance. The central thesis is that AI-driven value creation is not an inevitable outcome of technological adoption but rather a strategically mediated process contingent on responsible innovation practices, stakeholder engagement and robust governance frameworks. The study employs a systematic integrative literature review, synthesizing peer-reviewed journal articles, policy documents, industry reports and theoretical frameworks from Scopus, Web of Science, Google Scholar and authoritative bodies including the OECD, European Commission, United Nations and World Economic Forum. A thematic synthesis approach is used to identify patterns, tensions and gaps in the extant literature. The review reveals that AI creates economic, environmental and social value across sectors; however, this potential is systematically undermined by algorithmic bias, governance deficits and the marginalization of non-commercial stakeholders. Responsible innovation particularly through anticipation, reflexivity and inclusive engagement serves as a critical mediating mechanism between AI capabilities and sustainable outcomes. This paper integrates Stakeholder Theory, Responsible Innovation Theory and the Resource-Based View into an original conceptual framework that maps the pathways from AI adoption to sustainable value creation. It advances theoretical discourse by positioning responsible innovation as a dynamic capability and stakeholder engagement as a governance moderator. The framework offers actionable guidance for business executives, policymakers, AI designers and ESG strategists seeking to align technological innovation with sustainability imperatives and stakeholder expectations. Sustainable value creation in the AI era requires a deliberate integration of ethical, relational and governance considerations into the AI development and deployment lifecycle. Organizations that

embed responsible innovation principles and prioritize stakeholder perspectives are better positioned to achieve durable competitive advantage and societal legitimacy.

Keywords: AI, sustainable value creation, responsible innovation, ESG strategy, stakeholders.

1. Introduction

The rapid diffusion of artificial intelligence across global industries has fundamentally altered the landscape of organizational strategy, operations and governance. From autonomous decision systems to predictive analytics and natural language processing, AI represents a transformative technological wave whose socioeconomic implications remain incompletely understood (Brynjolfsson & McAfee, 2014; Russell & Norvig, 2020). Concurrently, the global sustainability agenda crystallized in the United Nations' (2015) Sustainable Development Goals calls for an integrated approach to economic development, social equity, and environmental stewardship. The convergence of these two forces raises a fundamental question: Can AI be a genuine driver of sustainable value, or does it risk deepening existing systemic inequalities and ecological burdens?

Traditional notions of corporate value centered on shareholder returns and short-term financial performance have been progressively challenged by broader conceptions of organizational purpose. Porter and Kramer (2011) introduced the concept of shared value, arguing that firms can simultaneously generate economic profit and address societal needs. Hart and Milstein (2003) extended this logic by proposing a sustainable value framework that integrates clean technology, pollution prevention, product stewardship, and future vision as pathways to long-term corporate sustainability. In an era increasingly shaped by ESG (Environmental, Social, and Governance) criteria and multi-stakeholder accountability, sustainable value creation has emerged as both a strategic imperative and a normative expectation (Friede et al., 2015).

Responsible innovation provides the normative and procedural architecture for ensuring that AI development serves societal good rather than narrow commercial interests. Defined by Stilgoe et al. (2013) as a transparent, interactive process in which societal actors and innovators become mutually responsive to one another, responsible innovation encompasses four dimensions: anticipation, reflexivity, inclusivity, and responsiveness. Applied to AI, this framework demands that organizations consider the potential social and environmental consequences of AI systems before, during, and after deployment not as an afterthought, but as a constitutive element of the innovation process itself.

The governance of AI has emerged as one of the defining challenges of the twenty-first century. Global bodies including the OECD (2019), the European Commission (2021), and the IEEE (2019) have issued normative frameworks emphasizing transparency, accountability, fairness, and human oversight. Yet translating these principles into organizational practice requires active stakeholder engagement. Freeman's (1984) stakeholder theory provides a foundational lens: organizations that systematically identify, engage, and balance the interests of diverse stakeholders employees, customers, communities, investors, and regulators are better positioned to develop AI systems that command social legitimacy and generate durable value.

Despite a growing body of scholarship on AI ethics (Jobin et al., 2019), AI-enabled sustainability (Vinuesa et al., 2020; Nishant et al., 2020) and responsible innovation (Owen et al., 2013), no integrative conceptual framework currently maps the multidimensional pathways through which AI capabilities, when governed by responsible innovation principles and stakeholder engagement,

translate into sustainable value creation. This paper addresses that gap. Specifically, the objectives are: (a) to synthesize the extant literature on AI, sustainability, responsible innovation, and stakeholder theory; (b) to identify moderating factors that shape the AI-sustainability relationship; and (c) to propose research directions that can empirically validate the proposed framework.

2. Conceptual and Theoretical Foundations

2.1 Artificial Intelligence and Digital Transformation

AI encompasses a family of computational technologies including machine learning, deep learning, computer vision, natural language processing, and reinforcement learning that enable machines to perform cognitive tasks traditionally associated with human intelligence (LeCun et al., 2015; Russell & Norvig, 2020). In organizational contexts, AI functions as a general-purpose technology (Brynjolfsson & McAfee, 2014) that reconfigures how firms create, deliver, and capture value. Duan et al. (2019) demonstrated that AI augments human decision-making quality and operational speed, while Agrawal et al. (2018) characterized AI primarily as a prediction technology that reduces uncertainty in complex decisions. Critically, AI does not operate autonomously from social context; its value-generating potential is mediated by organizational absorptive capacity, data quality, and governance structures.

2.2 Sustainable Value Creation

Sustainable value creation extends the classical economic concept of value to encompass environmental and social dimensions. Drawing on the triple bottom line framework (Elkington, 1997, as cited in Hart & Milstein, 2003), sustainable value creation requires organizations to generate economic returns without depleting natural capital or eroding social capital. Hart and Milstein's (2003) sustainable value framework identifies four strategic pathways: (1) reducing pollution and waste, (2) building tomorrow's competencies, (3) pursuing sustainable development, and (4) developing a vision for shared future prosperity. Porter and Kramer (2011) further operationalized this concept through shared value, arguing that competitive advantage and social progress are mutually reinforcing when organizations reconceive products, redefine productivity in the value chain, and build supportive industry clusters.

2.3 Responsible Innovation

The concept of responsible innovation emerged from science and technology studies to address the perceived disconnect between technical progress and broader societal benefit. Owen et al. (2012) argued that responsible innovation requires the scientific and business community to anticipate and reflect upon the wider implications and societal expectations of research and innovation. Stilgoe et al. (2013) operationalized this through the AREA framework -Anticipation, Reflection, Engagement and Action which provides a practical methodology for embedding societal considerations throughout the innovation process. Applied to AI, responsible innovation demands early identification of potential harms (anticipation), iterative reflexivity about design choices, inclusive engagement with affected communities, and adaptive responses to emerging ethical and social challenges.

2.4 Stakeholder Perspective

Freeman's (1984) seminal articulation of stakeholder theory fundamentally challenged the primacy of shareholder value in organizational decision-making. By defining stakeholders as any group or individual who can affect or is affected by the achievement of the organization's objectives, Freeman widened the scope of managerial accountability to include employees, customers, suppliers, communities, and the natural environment. Freeman et al. (2010) subsequently argued that the integration of stakeholder interests is not merely an ethical imperative but a source of competitive advantage firms that understand and address stakeholder needs more effectively reduce transaction costs, build reputational capital, and sustain innovation. In the AI context, stakeholder legitimacy is increasingly recognized as a prerequisite for AI adoption and societal acceptance (Ransbotham et al., 2020).

2.5 Theoretical Lens Used in the Paper

Stakeholder Theory

Stakeholder Theory provides the relational architecture for understanding how diverse groups perceive, respond to and are impacted by AI-driven organizational activities. It anchors the governance dimension of the framework by foregrounding accountability and stakeholder engagement as strategic imperatives.

Responsible Innovation Theory

Responsible Innovation Theory provides the normative and procedural logic for translating AI capabilities into outcomes that are ethically justifiable and socially legitimate. It serves as the central mediating mechanism in the proposed framework.

Resource-Based View / Dynamic Capabilities

Barney's (1991) Resource-Based View positions unique organizational resources and capabilities as the foundation of sustainable competitive advantage. Teece et al. (1997) extended this through the dynamic capabilities framework, emphasizing the firm's ability to integrate, build, and reconfigure competences in response to environmental change. In the AI context, data assets, algorithmic capabilities, and AI governance competencies constitute strategic resources whose value depends on how they are configured and deployed in accordance with sustainability and stakeholder principles (Teece, 2018).

3. Literature Review

3.1 AI and Organizational Value Creation

The relationship between AI adoption and organizational value creation has attracted substantial scholarly attention since the mid-2010s. Agrawal et al. (2018) argued that AI functions primarily as a prediction machine, lowering the cost of forecasting in business decisions and thereby creating value through reduced uncertainty and improved resource allocation. McKinsey Global Institute (Bughin et al., 2018) estimated that AI could generate approximately \$13 trillion in global economic activity by 2030, with the largest gains accruing in marketing, supply chain management, and manufacturing. Ransbotham et al. (2020), however, observed a persistent gap between AI ambition and realized value, attributing this to organizational learning deficits and

inadequate change management a finding that underscores the importance of dynamic capabilities in the AI value creation process.

3.2 AI and Sustainability Outcomes

The potential of AI to accelerate progress on global sustainability goals has been systematically documented by Vinuesa et al. (2020), who conducted a comprehensive analysis of the AI-SDG nexus and found that AI could enable 134 SDG targets across all 17 goals. In the environmental domain, AI applications in smart energy systems, precision agriculture, and climate modelling hold considerable promise for reducing resource consumption and carbon emissions (Nishant et al., 2020). Dauvergne (2020) introduced an important counterpoint, noting that AI's contribution to greening global supply chains is counterbalanced by the significant energy demands of training large AI models and the e-waste generated by rapid hardware obsolescence a tension that the sustainable AI discourse has begun to address in earnest.

3.3 Responsible AI: Ethics, Fairness, Transparency, Accountability

The ethics of AI has emerged as a vibrant interdisciplinary field. Jobin et al. (2019) conducted a comprehensive meta-analysis of 84 global AI ethics guidelines and identified 11 shared principles, with transparency, justice and fairness, and accountability appearing most frequently. Floridi et al. (2018) proposed a five-principle AI ethics framework beneficence, non-maleficence, autonomy, justice, and explicability which has been widely cited as a principlist foundation for AI governance. The European Commission (2021) operationalized these principles in its proposed AI Act, establishing a risk-based regulatory framework that requires high-risk AI systems to undergo conformity assessments, maintain transparency, and ensure human oversight. Zuboff (2019) offered a critical perspective, arguing that AI-enabled surveillance capitalism represents a systemic threat to human autonomy that ethical guidelines alone cannot adequately address.

3.4 Stakeholder Expectations and AI Legitimacy

As AI systems increasingly affect employment, access to credit, healthcare outcomes, and civic participation, the legitimacy of AI deployment has become a central concern for multiple stakeholder groups. Ransbotham et al. (2020) found that employee trust in AI recommendations is significantly influenced by perceptions of AI fairness and explainability. Freeman et al. (2010) argued that stakeholder engagement not only reduces organizational conflict but also generates the relational capital necessary for innovation. Organizations that fail to engage stakeholders in AI development risk encountering resistance, regulatory backlash and reputational damage outcomes that directly undermine sustainable value creation. The OECD (2019) principles on AI explicitly call for inclusive participation in AI governance as a prerequisite for trustworthy systems.

3.5 AI in ESG, Innovation, Operations and Decision-Making

AI's integration into ESG strategy represents a significant frontier in sustainable business practice. AI enables real-time monitoring of carbon emissions, supply chain ethics and governance compliance, providing organizations with granular ESG data that supports credible disclosure and strategy formulation (MSCI, 2022; World Economic Forum, 2018). In operations, AI-driven predictive maintenance, demand forecasting, and logistics optimization generate cost savings

while reducing material waste and energy consumption (Manyika et al., 2017). In decision-making, Duan et al. (2019) demonstrated that AI augments human cognitive capacities but also introduces new risks when algorithmic recommendations displace rather than support human judgment particularly in high-stakes social domains.

3.6 Gaps in the Existing Literature

A critical examination of the extant literature reveals several persistent gaps. First, studies on AI and sustainability tend to focus on specific sectors or technologies in isolation, without providing an integrative framework that links AI capabilities to multi-dimensional sustainability outcomes. Second, the responsible innovation literature has been slow to adapt its frameworks originally developed for upstream scientific research to the speed and scale of AI deployment in commercial organizations. Third, stakeholder theory has rarely been operationalized in the AI governance context in ways that move beyond normative prescription to actionable engagement frameworks. Fourth, empirical evidence on the moderating role of governance quality, data ethics, and organizational capabilities in the AI-sustainability relationship remains scarce.

4. Research Methodology

4.1 Research Design

This paper adopts a secondary-data-based conceptual and integrative literature review approach. Rather than generating new empirical data through primary methods, the study synthesizes and theorizes upon existing scholarly, policy, and industry knowledge. This approach is particularly appropriate for theoretical framework development (Owen et al., 2013) and for identifying gaps in emerging interdisciplinary fields where empirical evidence remains fragmented (Jobin et al., 2019).

4.2 Data Sources

The review drew upon literature from multiple credible databases and authoritative sources: academic databases including Scopus, Web of Science, SSRN, and Google Scholar; peer-reviewed journals in management, information systems, sustainability, and ethics; reports and policy documents from the OECD (2019), United Nations (2015), European Commission (2021), World Economic Forum (2018, 2020), McKinsey Global Institute, and IEEE (2019); and foundational theoretical texts from management and technology studies.

4.3 Search Strategy and Inclusion/Exclusion Criteria

The literature search employed Boolean operators combining terms such as 'artificial intelligence', 'sustainable value creation', 'responsible innovation', 'stakeholder theory', 'AI ethics', 'ESG', 'AI governance', and 'digital transformation'. Inclusion criteria encompassed peer-reviewed publications and authoritative reports published primarily between 2010 and 2024, written in English, and addressing at least one of the core constructs of the study. Seminal earlier works (e.g., Freeman, 1984; Barney, 1991; Teece et al., 1997) were included given their foundational theoretical relevance. Studies with narrow sectoral or geographic focus that lacked generalizability were noted but weighted accordingly.

5. Proposed Conceptual Framework

5.1 Core Logic of the Framework

The proposed framework illustrated in Figure 1 posits that the pathway from AI adoption to sustainable value creation is neither automatic nor linear. Rather, it is a mediated and moderated process in which responsible innovation serves as the central mechanism through which AI capabilities are converted into sustainable outcomes, while stakeholder engagement and governance structures function as critical contextual enablers and moderators. This architecture reflects the theoretical synthesis and it is consistent with the dynamic capabilities perspective that positions organizational processes not just resources as the source of durable advantage (Teece et al., 1997).

5.2 Role of Responsible Innovation

Within the framework, responsible innovation occupies the mediating position between AI capabilities and sustainable value creation. This positioning reflects Stilgoe et al.'s (2013) argument that the social and environmental outcomes of innovation are not determined by technical characteristics alone but by the governance and process architecture surrounding them. An organization that deploys AI with strong anticipatory practices (identifying long-term risks), reflective processes (questioning assumptions about AI design), inclusive engagement (involving diverse stakeholders), and adaptive responsiveness (adjusting systems based on feedback) is more likely to produce AI outcomes that are legitimate, equitable, and durable.

5.3 Role of Stakeholder Perspectives

Stakeholder perspectives function at two levels in the framework. First, they shape the demand environment within which AI systems operate stakeholders as customers, employees, and communities define the social acceptability and legitimacy of AI applications. Second, they function as governance participants, contributing to the co-design, oversight, and accountability mechanisms that ensure responsible AI deployment. Freeman et al. (2010) argued that firms that successfully engage stakeholders in value creation processes generate informational, relational, and reputational advantages that translate into superior long-term performance.

5.4 Pathways from AI Adoption to Sustainable Value Creation

The framework identifies three principal value pathways. The economic value pathway encompasses cost reduction, productivity enhancement, market intelligence, and revenue diversification enabled by AI-driven automation, prediction, and decision support (Bughin et al., 2018; Duan et al., 2019). The environmental value pathway leverages AI for energy optimization, emissions monitoring, waste reduction, and climate adaptation, thereby contributing to SDG 7, SDG 12, and SDG 13 (Vinuesa et al., 2020; Nishant et al., 2020). The social value pathway harnesses AI for healthcare access, educational equity, financial inclusion, and community resilience dimensions most directly threatened by algorithmic bias and digital exclusion (Floridi et al., 2018; Jobin et al., 2019).

5.5 Moderators and Enablers

The translation of AI capabilities into sustainable value is moderated by four contextual factors. First, governance and regulation the quality, consistency and enforcement of AI governance frameworks at organizational and institutional levels determines whether responsible innovation principles are operationalized or remain aspirational (European Commission, 2021; OECD, 2019). Second, organizational capabilities including AI literacy, data governance competencies, and cross-functional integration capacity determine whether firms can absorb and responsibly deploy AI technologies (Teece, 2018; Barney, 1991). Third, data quality encompassing completeness, representativeness, and ethical provenance directly shapes whether AI systems produce equitable or biased outcomes (Jobin et al., 2019). Fourth, ethical norms and culture the informal values, incentives, and behavioral expectations within organizations determine whether responsible innovation becomes embedded in practice or functions as a compliance exercise.

6. Discussion

The literature consistently affirms that AI possesses genuine transformative potential across all three dimensions of the triple bottom line. Economically, AI-enabled prediction, automation, and optimization generate measurable gains in productivity and market responsiveness (Agrawal et al., 2018; Manyika et al., 2017). Environmentally, applications in smart grid management, precision agriculture, and logistics optimization offer pathways to significant carbon reduction (Dauvergne, 2020; Vinuesa et al., 2020). Socially, AI can democratize access to expert knowledge in healthcare and education, improve disaster response coordination, and extend financial services to underserved populations (Nishant et al., 2020). The critical insight is that these value streams are not mutually exclusive indeed, organizations that pursue multi-dimensional value creation are more likely to generate the stakeholder legitimacy necessary for sustained AI adoption.

The business case for responsible AI is increasingly compelling, not solely on ethical grounds but on strategic ones. Irresponsible AI deployment characterized by opaque algorithms, biased outcomes, and inadequate oversight generates regulatory risk, reputational damage, and stakeholder antagonism that erodes both financial and social capital (Zuboff, 2019; World Economic Forum, 2020). Floridi et al. (2018) argued that ethical AI design is not a constraint on innovation but an enabler of trust and trust is the foundational infrastructure upon which long-term value is created. The parallel with the environmental sustainability discourse is instructive: just as firms that embedded environmental management early gained first-mover advantages in regulatory compliance and consumer preference, firms that embed responsible AI principles early are likely to build distinctive and durable competitive positions.

Trust is emerging as the defining currency of AI-driven value creation. Ransbotham et al. (2020) found that AI value realization is strongly correlated with organizational learning and employee engagement factors fundamentally shaped by perceptions of fairness and transparency. Freeman et al. (2010) argued that organizations earn legitimacy through demonstrated alignment between stated values and behavioral practices a standard that AI governance must meet. The OECD's (2019) recommendation that AI systems be transparent, explainable, and subject to human oversight reflects a broader societal demand that AI-generating organizations earn trust through accountability, not merely assert it through marketing.

The deployment of AI presents not only opportunities but also significant risks and trade-offs that responsible governance must navigate. The concentration of AI capabilities in a small number of

large technology firms raises antitrust and equity concerns that challenge the fair distribution of AI-generated value (Acemoglu & Restrepo, 2018). The energy intensity of training large language models and deep neural networks creates a tension between AI's potential environmental benefits and its own ecological footprint (Dauvergne, 2020). Algorithmic bias the tendency of AI systems trained on historically unequal data to perpetuate or amplify systemic discrimination represents a direct threat to the social dimension of sustainable value creation (Jobin et al., 2019). Zuboff (2019) identified surveillance capitalism the commodification of human behavioral data by AI-powered platforms as a structural risk to individual autonomy and democratic governance that transcends individual organizational practices.

Balancing the efficiency imperatives that drive AI adoption with the ethical and sustainability imperatives that condition its legitimacy requires organizational ambidexterity the capacity to exploit existing AI capabilities while simultaneously exploring responsible innovation pathways. Teece (2018) suggested that dynamic capabilities the ability to sense, seize, and reconfigure in response to changing environments provide the strategic architecture for this balance. Practically, this implies investing in AI ethics infrastructure (ethics boards, impact assessment tools, bias auditing); developing inclusive stakeholder engagement processes that go beyond regulatory compliance; integrating sustainability metrics into AI performance evaluation; and cultivating an organizational culture that normalizes reflexivity and accountability in AI development.

From a stakeholder theory perspective, the framework confirms that sustainable value creation is a fundamentally relational process one that requires organizations to manage not only technical systems but also the social contracts through which AI gains societal acceptance. From a responsible innovation perspective, the findings suggest that the AREA dimensions (Stilgoe et al., 2013) need to be operationalized as organizational routines embedded in AI development processes, not merely as ethical standards applied retrospectively. From a dynamic capabilities perspective, responsible AI governance can be understood as a higher-order organizational capability that enables firms to adapt their AI strategies in response to evolving ethical expectations, regulatory requirements, and stakeholder demands thereby constituting a source of sustainable competitive advantage in the Barnian (1991) sense.

7. Tables and Figures

Table 1

Definitions of Key Constructs from the Literature

Construct	Definition	Key Source(s)
Artificial Intelligence (AI)	The simulation of human cognitive processes by machines; encompasses machine learning, deep learning, natural language processing, and autonomous decision-making systems.	Russell & Norvig (2020); LeCun et al. (2015)
Sustainable Value Creation	The generation of economic, social, and environmental worth through business activities in ways that	Porter & Kramer (2011); Hart & Milstein (2003)

	preserve long-term stakeholder and ecological well-being.	
Responsible Innovation	A transparent, interactive process by which societal actors and innovators become mutually responsive to each other, aiming for ethically acceptable and sustainable innovation.	Stilgoe et al. (2013); Owen et al. (2013)
Stakeholder Theory	A managerial framework asserting that organizations create value for and are accountable to a broad set of stakeholders, not solely shareholders.	Freeman (1984); Freeman et al. (2010)
Responsible AI	The design, development, and deployment of AI systems that are fair, transparent, accountable, explainable, and aligned with human values and societal norms.	Jobin et al. (2019); IEEE (2019); European Commission (2021)
ESG (Environmental, Social, Governance)	A framework for evaluating how organizations manage environmental risks, social responsibilities, and governance structures in their operations and strategy.	Friede et al. (2015); MSCI (2022)
Dynamic Capabilities	The firm's capacity to integrate, build, and reconfigure internal and external competences to address rapidly changing environments.	Teece et al. (1997); Teece (2018)

Note. Constructs are defined as used in this paper. Sources represent primary definitional authorities; additional citations appear throughout the text.

Table 2

Summary of Major Studies on AI, Sustainability, and Responsible Innovation

Study	Domain	Key Focus	Finding/Insight	Limitation
Vinuesa et al. (2020)	SDGs & AI	AI as enabler/inhibitor of 17 SDGs	AI can positively influence 134 SDG targets; risks noted in 59	Lacks sector-specific analysis
Jobin et al. (2019)	AI Ethics	Global AI ethics guidelines convergence	11 principles identified; transparency & accountability most cited	Policy documents only, no organizational data
Stilgoe et al. (2013)	Responsible Innovation	AREA framework for RI	Anticipation, Reflection, Engagement,	Originally designed for R&D contexts, needs AI adaptation

			Action as core dimensions	
Floridi et al. (2018)	AI Governance	AI4People ethical framework	Five principles: beneficence, non-maleficence, autonomy, justice, explicability	Principlist; lacks implementation roadmap
Ransbotham et al. (2020)	Business AI	AI adoption and value creation in firms	Gap between AI awareness and tangible value realization persists	Survey-based; limited longitudinal depth
Duan et al. (2019)	AI & Decision-making	AI in intelligent decision support	AI augments human decision quality and speed; organizational context critical	Primarily Western firm sample
Makridakis (2017)	AI & Economy	Future AI impact on economy and society	AI will radically alter labor markets; requires proactive governance	Speculative; limited empirical base
Nishant et al. (2020)	IS & Sustainability	AI for environmental sustainability	AI aligns with triple-bottom-line goals; data access is a barrier	Narrow focus; excludes social sustainability

Table 3

Thematic Synthesis: Opportunities, Risks, and Sustainability Outcomes

Theme	Opportunities	Risks / Trade-offs	Sustainability Outcomes
AI & Economic Value	Cost reduction, operational efficiency, revenue growth, process automation	Job displacement, market concentration, income inequality	GDP growth, productivity gains, SME competitiveness
AI & Environmental Value	Energy optimization, smart grids, precision agriculture, climate modelling	High energy consumption of AI models; e-waste from hardware	Carbon footprint reduction, SDG 7 (Clean Energy), SDG 13 (Climate Action)

AI & Social Value	Healthcare diagnostics, inclusive finance, educational access, disaster response	Algorithmic bias, digital divide, surveillance risks, privacy erosion	Improved health outcomes, social equity, SDG 3, SDG 10
Responsible AI Governance	Regulatory frameworks, ethics boards, third-party auditing, explainability tools	Compliance costs, innovation slowdown, fragmented global standards	Trust, accountability, legitimacy in AI deployment
Stakeholder Engagement	Co-design with communities, inclusive AI development, feedback mechanisms	Stakeholder conflicts, consultation fatigue, power asymmetries	Increased adoption, social licence, brand equity

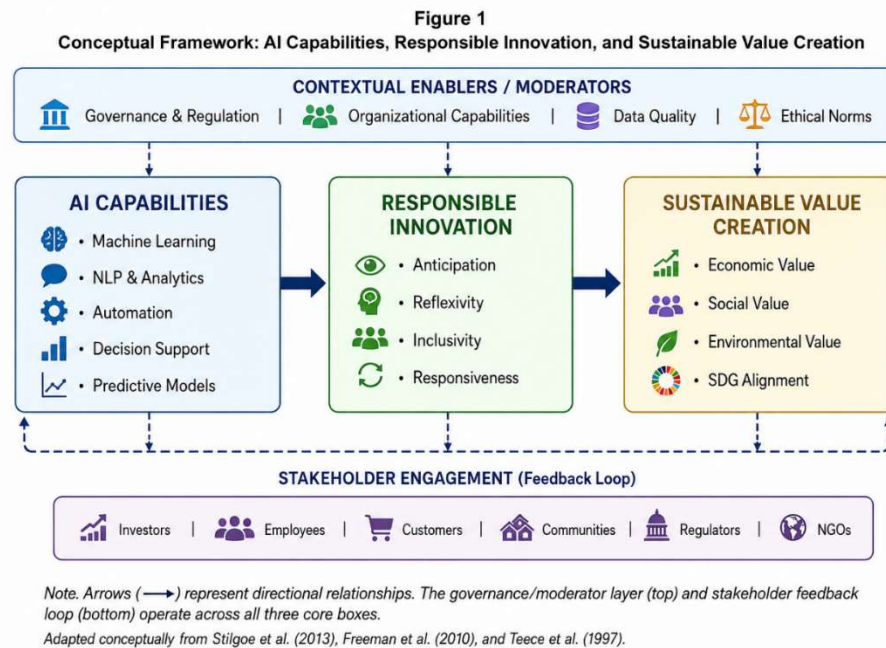
Note. Themes derived from integrative review of sources cited in the literature review. Trade-offs represent areas requiring active governance and stakeholder management.

Table 4

Future Research Agenda: Unresolved Gaps and Suggested Directions

Research Gap	Suggested Research Question	Recommended Method	Theoretical Lens
AI-sustainability link in emerging markets	How does AI adoption influence sustainable value creation in developing economies?	Longitudinal case study; cross-country survey	Stakeholder Theory, RBV
Algorithmic bias & social equity	How do biased AI systems undermine ESG social scores and stakeholder trust?	Mixed-methods; audit studies	Responsible Innovation, Institutional Theory
Governance effectiveness of AI ethics frameworks	Which AI governance models produce the best responsible innovation outcomes?	Comparative policy analysis; panel data	Regulatory theory, dynamic capabilities
Stakeholder perceptions of AI legitimacy	How do different stakeholder groups (employees, communities, investors) assess AI legitimacy?	Survey-based; qualitative interviews	Legitimacy Theory, Stakeholder Theory
AI and long-term ESG performance	Does responsible AI adoption causally improve firm-level ESG ratings over time?	Longitudinal panel study; causal inference models	RBV, Sustainability Theory

Figure 1



Note. Arrows (→) represent directional relationships. The governance/moderator layer (top) and stakeholder feedback loop (bottom) operate across all three core boxes. Adapted conceptually from Stilgoe et al. (2013), Freeman et al. (2010), and Teece et al. (1997).

8. Managerial and Policy Implications

For chief executives and boards, the framework suggests that AI strategy must be reframed from a technology investment question to a value creation and governance question. Leaders should institutionalize responsible AI principles at the board level, appoint dedicated AI ethics officers, and integrate sustainability outcomes into AI performance metrics alongside financial KPIs. Ransbotham et al. (2020) emphasized that the firms extracting greatest value from AI are those that invest simultaneously in technical capability and organizational learning — a finding that underscores the strategic importance of responsible innovation as a dynamic capability.

Policymakers should develop risk-proportionate regulatory frameworks that foster responsible AI innovation without stifling competitive dynamism. The European Commission's (2021) AI Act provides a useful template, though its implementation demands complementary investment in regulatory capacity particularly in developing economies where AI governance institutions are nascent. The OECD (2019) recommendations provide a globally applicable baseline that national regulators can adapt to local legal and cultural contexts. International regulatory cooperation is essential to prevent regulatory arbitrage, where firms domicile AI development in low-governance jurisdictions to evade accountability.

AI designers and technology managers bear primary responsibility for operationalizing responsible innovation principles at the technical level. This requires designing for explicability and transparency from the outset, not as a post-hoc addition; conducting mandatory bias audits on training data and model outputs; embedding privacy-by-design principles consistent with emerging data protection regulations; and developing mechanisms for ongoing stakeholder

feedback that allow AI systems to be iteratively improved in response to real-world impacts (Floridi et al., 2018; IEEE, 2019).

AI offers ESG practitioners powerful tools for data collection, reporting, and strategy execution. Organizations can deploy AI to monitor supply chain emissions in near real time, identify governance anomalies in financial reporting, and assess the social impacts of operations with greater granularity than traditional survey methods allow. However, ESG strategists must guard against the temptation to use AI as a greenwashing instrument employing sophisticated analytics to create the appearance of sustainability performance without substantive underlying change. Friede et al. (2015) demonstrated that genuine ESG performance generates long-term financial outperformance; AI should be deployed in service of this genuine performance, not its simulation.

Organizations must transition from viewing stakeholder engagement as a compliance function to recognizing it as a strategic input into AI development and governance. This means establishing formal multi-stakeholder advisory bodies for AI oversight; creating accessible channels through which customers, employees, and communities can report AI-related harms or concerns; and disclosing AI systems' intended purposes, capabilities, and limitations in plain language. Freeman's (1984) insight that identifying and managing stakeholder interests is a source of competitive advantage is nowhere more applicable than in the governance of AI, where trust deficits can rapidly translate into adoption barriers and regulatory escalation.

9. Future Research Directions

The conceptual framework presented in this paper requires empirical validation across multiple levels of analysis. At the firm level, quantitative studies could employ structural equation modelling to test the mediation of responsible innovation between AI capability and sustainability performance, using ESG scores as proxy dependent variables. At the industry level, comparative studies could examine the extent to which different AI governance architectures (regulatory, voluntary, co-regulatory) predict responsible innovation outcomes. Experimental approaches could test stakeholder responses to different AI transparency disclosure formats.

The framework's generalizability should be tested across diverse organizational, sectoral, and national contexts. High-priority sectors include healthcare (where AI decisions directly affect patient welfare), financial services (where algorithmic bias has well-documented effects on credit access), and public administration (where AI governs citizen-facing services and requires democratic accountability). Each sector presents distinct stakeholder configurations and governance challenges that will enrich and potentially refine the framework. The dynamic and path-dependent nature of AI-driven value creation demands longitudinal research designs that can capture how responsible innovation practices and stakeholder governance evolve over the AI adoption lifecycle. Mixed-method studies combining quantitative ESG and financial data with qualitative case study insights would provide the methodological triangulation necessary to establish causal mechanisms rather than merely associational relationships.

10. Limitations

This paper is subject to several limitations inherent in its methodological design. First, as a conceptual framework paper based on secondary data, it does not provide empirical validation of the proposed relationships. The propositions advanced remain theoretical until subjected to

rigorous empirical testing across diverse organizational and national contexts. Second, the integrative review methodology, while systematic in design, involves interpretive judgements in the selection and synthesis of studies judgements that, despite the author's efforts at transparency and rigor, are inevitably subject to some degree of subjectivity. Third, the literature reviewed skews toward English-language publications from Western academic traditions, which may underrepresent important insights from non-Western AI governance practices and sustainability discourses. Fourth, the rapidly evolving nature of AI technology means that some studies reviewed may already be partially superseded by subsequent technical developments a challenge inherent in any scholarly engagement with a field advancing at the pace of contemporary AI. These limitations are not unique to the present paper but are characteristic of the genre; they are best addressed through the empirical and cross-contextual future research directions.

11. Conclusion

This paper has contributed to a new perspective identifying AI-driven sustainable value creation as a strategically mediated process and not as a natural consequence of technology. On the basis of an in-depth integrative literature review and the synthesis of three streams of theory (the Stakeholder Theory, Responsible Innovation Theory and the Resource-Based View), it proposes an original conceptual framework that examines the pathways of AI capabilities to sustainable economic, social and environmental value, mediated by responsible innovation, and moderated by governance quality, organizational capabilities, data ethics and stakeholder engagement. There are three theoretical contributions of the framework. It offers, first of all, the first theoretical architecture that is integrative and formally includes responsible innovation as a mediating element in the relationship between AI and sustainability. Secondly, it moves stakeholder engagement from a compliance task to an active governance tool to influence AI's legitimacy and social licence. Thirdly, it includes the dynamic capabilities logic, allowing for a resource-based perspective on why responsible AI governance is not just an ethical duty but also a source of sustainable competitive advantage. Another very important application impact is the one that involves application impact. It provides a blueprint for business leaders on how to incorporate sustainability and ethical governance into their AI strategies.

References

- Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 197–236). University of Chicago Press.
- Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Review Press.
- Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120. <https://doi.org/10.1177/014920639101700108>
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton.
- Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Notes from the AI frontier: Modeling the impact of AI on the world economy*. McKinsey Global Institute.

- Dauvergne, P. (2020). Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. *Review of International Studies*, 47(1), 81–97. <https://doi.org/10.1017/S0260210520000078>
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63–71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
- European Commission. (2021). Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). European Commission. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
- 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). 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.
- Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., & de Colle, S. (2010). *Stakeholder theory: The state of the art*. Cambridge University Press.
- Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), 210–233. <https://doi.org/10.1080/20430795.2015.1118917>
- Hart, S. L., & Milstein, M. B. (2003). Creating sustainable value. *Academy of Management Perspectives*, 17(2), 56–67. <https://doi.org/10.5465/ame.2003.10025194>
- IEEE. (2019). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems (1st ed.)*. IEEE. <https://ethicsinaction.ieee.org>
- 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>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Liang, H., Zuo, M., & Hu, Q. (2023). Artificial intelligence, digital transformation, and corporate sustainability. *Sustainability*, 15(8), 6619. <https://doi.org/10.3390/su15086619>
- Makridakis, S. (2017). The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46–60. <https://doi.org/10.1016/j.futures.2017.03.006>
- Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). *A future that works: Automation, employment, and productivity*. McKinsey Global Institute.
- MSCI. (2022). *MSCI ESG ratings methodology*. MSCI ESG Research LLC. <https://www.msci.com/documents/1296102/21901542/MSCI+ESG+Ratings+Methodology+-+Exec+Summary+2022.pdf>

- Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management*, 53, 102–104. <https://doi.org/10.1016/j.ijinfomgt.2020.102104>
- OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy*, 39(6), 751–760. <https://doi.org/10.1093/scipol/scs093>
- Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A framework for responsible innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible innovation: Managing the responsible emergence of science and innovation in society* (pp. 27–50). Wiley.
- Porter, M. E., & Kramer, M. R. (2011). Creating shared value. *Harvard Business Review*, 89(1/2), 62–77.
- Ransbotham, S., Khodabandeh, S., Kiron, D., Candelon, F., Chu, M., & LaFountain, B. (2020). Expanding AI's impact with organizational learning. MIT Sloan Management Review and Boston Consulting Group. <https://sloanreview.mit.edu/projects/expanding-ais-impact-with-organizational-learning/>
- Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580. <https://doi.org/10.1016/j.respol.2013.05.008>
- Teece, D. J. (2018). Business models and dynamic capabilities. *Long Range Planning*, 51(1), 40–49. <https://doi.org/10.1016/j.lrp.2017.06.007>
- 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)
- United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development (A/RES/70/1). United Nations. <https://sdgs.un.org/2030agenda>
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Fellander, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1), 233. <https://doi.org/10.1038/s41467-019-14108-y>
- World Economic Forum. (2018). The new physics of financial services: How artificial intelligence is transforming the financial ecosystem. WEF. <https://www.weforum.org/reports/the-new-physics-of-financial-services>
- World Economic Forum. (2020). The global risks report 2020 (15th ed.). WEF. <https://www.weforum.org/reports/the-global-risks-report-2020>
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.