American Journal of Applied Research and AI (AJARAI)

VOLUME 1 ISSUE 1 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of
Applied Research and AI (AJARAI))

Volume 1 Issue 1, Year 2026
https://journals.e-palli.com/home/index.php/ajarai

Machine Learning in Corporate Financial Sustainability: A Critical Evaluation of Models Bias and Outcomes

Zulkiffly Baharom[1*]

Article Information

Received: November 09, 2025
Accepted: January 29, 2026
Published: March 26, 2026

Keywords

Algorithmic Bias, Corporate Financial Sustainability, ESG, Explainable Machine Learning, Machine Learning

ABSTRACT

The integration of machine learning (ML) into corporate financial sustainability (CFS) is a double-edged sword: while offering transformative potential for predictive analytics, risk modeling, and reporting efficiency, it also introduces significant risks of algorithmic bias, opacity, and erosion of accountability. This critical literature review synthesizes 31 peer-reviewed articles to evaluate ML's role in CFS contexts, moving beyond techno-optimism to foreground ethical and governance challenges. Our analysis reveals that current applications, such as ESG scoring, predictive CFS analytics, and automated reporting, often prioritize scalability and efficiency over validity, equity, and substantive performance, thereby risking "automated greenwashing" and reinforcing structural inequalities. In response, we propose an integrative, four-pillar framework for responsible ML deployment, emphasizing Transparent and Explainable Models, Bias-Audit and Ethical Governance, Integrated ESG-ML Reporting Standards, and Stakeholder-Inclusive ML Deployment. Institutional, technological, and cultural contexts moderate the effectiveness of these pillars. We argue that without deliberate governance, ML may undermine the very goals of sustainable value creation it seeks to advance. This review calls for interdisciplinary collaboration, standardized auditing protocols, and proactive regulation to align ML innovation with the long-term imperatives of CFS.

INTRODUCTION

The convergence of ML and the global imperative for CFS marks a pivotal and contentious frontier in modern business governance. As ML algorithms become deeply embedded in strategic decision-making, risk assessment, and performance measurement, their application to environmental, social, and governance (ESG) objectives and CFS has drawn significant attention from both scholars and practitioners (Shabbir, 2025; Zhao & Gómez Fariñas, 2023). Proponents contend that ML-driven analytics can unlock unprecedented efficiencies, enhance the precision of sustainability metrics, and strengthen long-term corporate resilience against systemic shocks (Fang *et al.*, 2025; Li *et al.*, 2025). However, this technological optimism is tempered by escalating concerns about algorithmic opacity, embedded biases, and the potential for ML to automate and amplify existing governance failures (Blue *et al.*, 2025; Seele, 2017).

Emerging literature at this intersection remains fragmented, often oscillating between uncritical techno-optimism and abstract ethical caution without offering integrative, governance-oriented frameworks that reconcile efficiency with equity. This critical review addresses that gap by systematically examining how ML models are conceptualized, applied, and governed within the domain of CFS, defined here as a firm's capacity to maintain economic viability, solvency, and adaptive strength over the long term, a condition increasingly understood as interdependent with robust ESG performance and stakeholder trust (Sandberg *et al.*, 2023; Sulimany *et al.*, 2021).

Guided by three central research questions, this article moves beyond a descriptive summary to offer a constructive critique and a forward-looking governance model:

1.How are ML models currently operationalized in CFS contexts, and what documented or potential outcomes are observed for performance, reporting, and risk management?

2.What specific forms of bias, ethical challenges, and governance gaps arise in the development and deployment of ML-driven CFS systems?

3.How can corporate governance structures and regulatory frameworks be redesigned to ensure that ML supports authentic, equitable, and long-term sustainable value creation rather than short-term optimization or "sustainability-washing"?

To address these questions, we employ a Critical Literature Review (CLR) methodology, analyzing 31 peer-reviewed articles to synthesize dispersed insights, highlight critical tensions, and propose a pathway for the responsible integration of ML. In response to the identified governance deficit, we introduce an integrative, four-pillar framework for responsible ML deployment, encompassing Transparent and Explainable Models, Bias-Audit and Ethical Governance, Integrated ESG-ML Reporting Standards, and Stakeholder-Inclusive ML Deployment. This framework is designed to align technological capability with the normative goals of sustainable corporate governance, and institutional, technological, and cultural contexts moderate its effectiveness.

[1] Tunku Puteri Intan Safinaz School of Accountancy (TISSA-UUM), College of Business, Universiti Utara Malaysia, Malaysia
[*] Corresponding author's email: zulkifflybaharom100@gmail.com

By foregrounding the tensions between instrumental efficiency and ethical governance, this review aims to inform scholars, practitioners, and policymakers seeking to harness ML not merely as an optimization tool but as a steward of long-term, equitable, and resilient value creation.

LITERATURE REVIEW

A coherent understanding of the ML-CFS nexus requires grounding in its core constructs and theoretical lenses, as well as a critical examination of current applications. This section defines key terms, outlines the theoretical foundations, and, through a critical lens, synthesizes how ML is currently applied in CFS contexts.

Defining Core Constructs

ML in CFS refers to the application of ML techniques, including supervised learning, unsupervised clustering, natural language processing (NLP), and neural networks, to analyze large-scale, often unstructured datasets for tasks central to CFS management. These tasks include ESG rating generation, carbon emission forecasting, detection of misleading disclosure ("greenwashing"), predictive modeling of climate-related financial risks, and optimization of social impact metrics (Ferro *et al.*, 2025; Hąbek, 2025; Subramaniam *et al.*, 2024).

CFS differs from broader corporate sustainability in its focus on a firm's enduring economic health. It encompasses long-term liquidity, solvency, profitability, and, critically, resilience, defined as the capacity to absorb, adapt to, and recover from systemic disruptions (Knoepfel, 2001). Contemporary scholarship increasingly argues that CFS is inextricably linked to ESG performance, as environmental liabilities, social license to operate, and governance failures constitute material financial risks (Allen *et al.*, 2024; Sandberg *et al.*, 2023).

Model Bias and Explainability are two interrelated ethical and technical challenges in ML deployment. Algorithmic bias refers to systematic, unfair errors in model outputs that disadvantage certain groups or lead to biased outcomes, often stemming from unrepresentative training data, flawed problem formulation, or biased feature selection (Downing, 2025). Explainability (or interpretability) is the degree to which human stakeholders can understand a model's internal logic and decision-making process, a necessity for accountability, audit, and trust, mainly when complex "black-box" models like deep neural networks are employed (Schneider & Schwab, 2025; Suárez Giri & Sanchez-Chaparro, 2024).

Theoretical Foundations

Four theoretical perspectives provide essential grounding for analyzing the ML-CFS interface.

Resource-Based View (RBV)

From an RBV perspective, ML capabilities are strategic, intangible assets that can be a source of sustained competitive advantage if they are valuable, rare, inimitable, and non-substitutable (Lin *et al.*, 2006). ML-driven CFS analytics can thus be framed as a dynamic capability that enhances a firm's ability to identify ESG risks and opportunities, innovate sustainably, and build reputational capital (Li *et al.*, 2025; Martin *et al.*, 2016).

Stakeholder Theory

This theory holds that firms manage relationships with a broad constellation of stakeholders whose interests are intrinsically valuable (Freeman, 1984). ML systems deployed in CFS contexts must therefore be evaluated not only for technical efficiency but also for their impact on stakeholders, including employees subject to algorithmic monitoring, communities affected by environmental predictions, and investors relying on ML-generated ESG scores (Atif *et al.*, 2023; Strätling, 2007). A lack of transparency or fairness can erode stakeholder trust and legitimacy.

Institutional Theory

Institutional pressures, coercive (regulation), mimetic (industry benchmarking), and normative (professional standards), powerfully shape corporate behavior (DiMaggio & Powell, 1983). The rapid adoption of ML for CFS reporting can be interpreted as a response to regulatory mandates, such as the EU's Corporate Sustainability Reporting Directive (CSRD), investor demand for ESG data, and industry norms regarding digitalization (Gribnau, 2024; Knoepfel, 2001).

Ethical ML Governance

This emerging interdisciplinary field provides normative principles, such as fairness, accountability, transparency, and privacy (FAT-P), for designing and governing ML systems (Jobin *et al.*, 2019). It requires that sustainability-focused ML move beyond a purely instrumental logic to incorporate ethical safeguards, human oversight, and mechanisms for redress (Zhao & Gómez Fariñas, 2023).

While these theories offer valuable lenses, they also have limitations in the ML-CFS context. For instance, RBV may overlook negative externalities and ethical trade-offs; Stakeholder Theory may struggle to operationalize multi-stakeholder inputs in algorithmic design; and Institutional Theory may not fully capture the pace of technological change, which is outpacing regulatory adaptation. Acknowledging these limitations sets the stage for the critical analysis that follows (Table 1).

Table 1: Critical Analysis of ML Application

Application Area	Key ML Techniques	Promised Benefits	Critical Risks/Limitations	Key References
ESG Scoring & Ratings	NLP, Supervised Learning	Scalability, consistency, timeliness	Reliance on aspirational data; opacity; amplification of bias	Ferro *et al.* (2025); Suárez Giri & Sánchez-Chaparro (2024)

Predictive Sustainability Analytics	Regression, Neural Networks	Proactive management, target setting	"Green Paradox" of AI; data quality issues; metric myopia	Fang *et al.* (2025); Hsieh (2024)
Automated Reporting & Assurance	NLP, Data Integration Systems	Efficiency, reduced human error	Conflation of efficiency with quality; assurance challenges	Hąbek (2025); Naveed *et al.* (2025)
Risk Resilience Modelling	Simulation, Scenario Analysis	Enhanced foresight, systemic risk insight	Governance readiness gap; ethical concerns of predictive control	Knoepfel (2001); Seele (2017)
Fraud & Anomaly Detection	Anomaly Detection, Pattern Recognition	Integrity monitoring, fraud prevention	Arms race with evasion techniques; over-reliance on historical data	Blue *et al.* (2025)

Current ML Applications in CFS: A Critical Analysis

The literature shows a rapidly expanding portfolio of ML applications in CFS, promising improved efficiency, accuracy, and insight. However, a critical analysis reveals that each application area is fraught with methodological, ethical, and substantive limitations that can undermine its contribution to genuine CFS.

ESG Scoring and Ratings

The use of ML, particularly NLP, to generate ESG scores is among the most widespread applications. Algorithms parse unstructured data from CFS reports, news, and filings to produce standardized ratings (Ferro *et al.*, 2025). While this enables scalable assessment, critical analysis exposes serious flaws. Ferro *et al.* (2025) found that approximately 60% of a major provider's ESG score relied on aspirational rather than performance data, effectively measuring rhetoric over reality, a form of "automated greenwashing." Furthermore, the proprietary nature of these models creates a "black box within a black box" (Suárez Giri & Sanchez-Chaparro, 2024), obscuring accountability and disadvantaging firms with fewer disclosure resources (Downing, 2025).

Predictive Sustainability Analytics

ML models are increasingly used to forecast metrics such as carbon emissions and resource use (Fang *et al.*, 2025). While enabling proactive management, these systems face the "Green Paradox of ML," in which short-term costs and disruptions can depress performance, deterring investment. Moreover, prediction accuracy depends on data quality, which is often lacking and may foster "metric myopia", overemphasizing quantifiable indicators at the expense of harder-to-measure social and ethical dimensions.

Automated Reporting and Assurance

ML-driven automation of CFS data collection and report generation is promoted to ease compliance with standards such as the CSRD (Hąbek, 2025). However, this risks conflating reporting efficiency with reporting quality. The inability to audit the underlying ML models undermines traditional assurance mechanisms and may erode stakeholder trust in automated disclosures (Naveed *et al.*, 2025).

Risk Resilience Modeling

ML offers sophisticated tools for modeling climate-related and systemic risks (Seele, 2017). However, a governance readiness gap often persists: boardrooms may lack the literacy to interpret probabilistic outputs, leading to neglect or overconfidence. Ethically, such predictive systems raise concerns about surveillance, preemptive intervention, and the marginalization of nonquantifiable resilience factors such as organizational culture and stakeholder relationships.

Creative Accounting and Fraud Detection

ML shows promise in detecting patterns of financial misrepresentation (Blue *et al.*, 2025). However, this is inherently an arms race, with evasion tactics evolving alongside detection tools. Over-reliance on historical data may also limit the identification of novel fraud schemes, and the shift toward algorithmic auditing may reduce the role of contextual human judgment.

In summary, while ML applications in CFS are diverse and expanding, each entails critical trade-offs between scalability and validity, short-term cost and long-term gain, automated efficiency and human accountability, and predictive quantification and holistic resilience. Recognizing these tensions is essential to developing governance frameworks that ensure ML delivers genuine sustainable value rather than becoming a vector for obfuscation, bias, or risk.

MATERIAL AND METHODS

This study employs a CLR methodology, selected for its capacity to provide a deep, analytical, and interpretive synthesis of a complex and contested body of knowledge (Adams & Whelan, 2009). Unlike systematic reviews focused on quantitative aggregation, a CLR critically engages with the underlying arguments, assumptions, intellectual evolution, and ideological currents within a field. The objective is to move beyond cataloging findings to interrogate conceptual coherence, reveal tensions, and identify pathways for future inquiry.

The article selection process began with a structured search of the Scopus database, chosen for its high-quality, peer-reviewed coverage of business, economics,

and management journals. The search strategy combined keywords across three domains:

• ML/AI: ("machine learning," "artificial intelligence," "predictive analytics," "algorithm*")

• Financial Sustainability: ("financial sustainability," "long-term financial performance," "corporate resilience")

• Core Challenges: ("bias," "transparency," "explainability," "ethical AI," "black box")

Filters were applied to include English-language articles published between 2017 and 2025, with a focus on final-stage articles and review papers. The year 2017 was chosen as a meaningful starting point because of the emergence of seminal works linking ML governance and CFS (e.g., Seele, 2017).

The initial search yielded 158 articles. A rigorous screening process followed, guided by explicit inclusion and exclusion criteria:

Included

Articles directly addressing conceptual, empirical, or critical links between ML/AI and CFS or ESG performance.

Excluded

Studies focusing solely on technical ML model development without governance or CFS implications, or those set in purely non-corporate contexts.

This process refined the corpus to 31 articles for in-depth analysis.

Critical Analysis Procedure

The analysis employed thematic analysis to identify recurring patterns in how ML applications, benefits, and risks were framed across the literature. Argument analysis was used to deconstruct central claims and supporting evidence, mapping areas of consensus and debate. Particular attention was paid to the following:

Theory–Evidence Alignment

Examining whether empirical findings substantiate theoretical claims.

Conceptual Gaps

Identifying conflation (e.g., between ML adoption and CFS achievement) and under-explored areas (e.g., longitudinal causal studies).

Ideological Positioning

Noting whether articles leaned toward techno-solutionism, ethical caution, or integrative governance perspectives.

To ensure thematic saturation, the analysis continued until no new themes emerged from the dataset, and conflicting viewpoints were systematically compared to avoid bias toward a single narrative.

RESULTS AND DISCUSSION

Results

The critical analysis of the 31-article corpus reveals a field marked by promising applications, significant documented risks, and moderating factors that shape outcomes.

ML Models in CFS: Applications and Outcomes

The literature shows a rapid expansion of ML applications with mixed and context-dependent outcomes. A substantial body of research highlights gains in efficiency and accuracy. For instance, ML models outperform traditional methods in handling missing ESG data, producing more robust scores (Downing, 2025). Neural networks that incorporate textual data from news articles have been shown to predict firm financial ratios more accurately than models relying solely on historical numerical data (Zhai & Zhang, 2023). Furthermore, firm-level ML adoption is associated with improved ESG disclosure quality, particularly in the governance and social dimensions, through mechanisms such as enhanced transparency and talent recruitment (Zhou *et al.*, 2025; Naveed *et al.*, 2025).

However, these benefits are tempered by significant limitations. The literature confirms the "green paradox of ML/AI" (Fang *et al.*, 2025), in which short-term integration costs and operational disruptions can initially pressure financial performance before long-term benefits materialize. More critically, studies question the substantive validity of ML-enabled metrics. Ferro *et al.* (2025) found that approximately 60% of a major provider's ESG score was based on forward-looking promises rather than verified performance data. This suggests that ML may efficiently process corporate rhetoric without verifying the reality of CFS, thereby sophisticating presentation over performance.

Identified Biases and Ethical Challenges

The review highlights various ongoing and harmful biases in ML systems for CFS.

Data Bias and Structural Advantage

A pervasive issue is "missing data bias," in which firms with greater resources (typically larger, developed-market companies) produce more comprehensive CFS disclosures. ML models trained on this incomplete data systematically favor these firms, penalizing small and medium-sized enterprises (SMEs) and those in emerging economies (Downing, 2025; Bernardini *et al.*, 2024).

Algorithmic Opacity and the "Black Box"

The inherent opacity of advanced ML models, especially deep learning architectures, undermines accountability in high-stakes sustainability contexts. When an ML system denies a green loan or flags a firm for human rights risks, the inability to explain "why" hampers stakeholder engagement and complicates regulatory oversight (Suárez Giri & Sanchez-Chaparro, 2024; Shilbayeh & Grassa, 2024). Alipasa *et al.* (2026) also observed that trust in machine-learning-supported evaluation remains

moderate, with faculty showing greater confidence than students. This underscores the need for transparency and human oversight when ethically integrating automated systems into high-stakes environments.

Normative and Measurement Bias

Significant divergence in ESG ratings for the same company stems from providers embedding subjective weightings in proprietary algorithms (Cheng *et al.*, 2025). This lack of standardization can amplify conflicting notions of "good" CFS, confusing investors and managers.

Surveillance and Privacy Risks

Using ML to monitor supply chain labor practices or employee behavior, often framed as social governance (S), raises serious ethical concerns about worker surveillance, data privacy, and consent (Hsieh, 2024; Seele, 2017).

Moderating Factors Influencing Outcomes

The relationship between ML adoption and CFS outcomes is not deterministic but is shaped by key moderating variables:

Regulatory and Institutional Context

Stringent regulations (e.g., CSRD) drive more substantive ML integration for compliance, whereas weaker regimes may foster superficial or biased applications (Gribnau, 2024).

Industry and Firm Characteristics

ML's impact is more pronounced in high-ESG-materiality sectors (e.g., energy, manufacturing) and in firms with preexisting technological maturity (Khan & Gupta, 2025; Li *et al.*, 2025).

Governance Structures

The presence of dedicated CFS committees strengthens the positive relationship between ML adoption and disclosure quality, serving as a crucial governance mediator (Naveed *et al.*, 2025; Vintilă *et al.*, 2025).

National and Cultural Context

Effects vary across developed and emerging markets and between state-owned and private enterprises, indicating that ownership structures, market maturity, and cultural attitudes toward technology and transparency are critical moderators (Spagnuolo *et al.*, 2025; Allen *et al.*, 2024).

Discussion

The findings of this critical review illuminate a central, unresolved dialectic in the ML-CFS nexus: the tension between instrumental efficiency and normative governance. On the one hand, ML is championed as a powerful tool for optimizing processes, reducing costs, and generating precise predictions, objectives aligned with traditional, efficiency-centric models of corporate performance. On the other hand, CFS is fundamentally a normative endeavor concerned with equity, intergenerational justice, long-term stewardship, and multi-stakeholder accountability. The uncritical application of ML, optimized for narrow financial or operational metrics, risks subverting these normative goals by automating bias, obscuring accountability, and privileging measurable outputs over meaningful outcomes.

Theoretical Implications and Connections to Broader Debates

This tension echoes enduring debates in corporate governance, particularly the conflict between shareholder primacy and stakeholder theory. The "black box" of ML introduces a new form of agency problem, in which managers or technologists deploy inscrutable systems that may serve narrow interests, such as short-term financial performance or streamlined compliance, at the expense of long-term resilience and stakeholder trust. The call for explainable ML (XML) and algorithmic auditing thus parallels historical demands for transparent financial reporting and independent board oversight.

Furthermore, the findings challenge the Resource-Based View (RBV) by showing that ML capabilities, while potentially valuable and rare, can also pose risks and illegitimacy if not governed ethically. This suggests the need for an expanded RBV that treats ethical and reputational capital as critical components of sustained advantage.

From an Institutional Theory perspective, the rapid but uneven adoption of ML for CFS reflects not only coercive regulatory pressures but also mimetic isomorphism in the absence of standards. The divergence in ESG ratings and the "green paradox" reveal how institutional voids can lead to symbolic rather than substantive adoption of ML, undermining its potential for genuine impact.

Practical Implications for Stakeholders

For Corporate Boards and Executives

ML governance must become a core board competency, moving beyond procurement to oversee model design, bias testing, and impact assessment. Boards should:

- Establish ML ethics committees with cross-disciplinary expertise.
- Mandate regular algorithmic audits and transparent disclosure of ML use cases in CFS reporting.
- Develop ML literacy programs for directors to enable effective oversight of complex models.

For Regulators and Standard-Setters

There is a pressing need to develop frameworks not only for CFS disclosure but also for disclosing ML's role in preparing those disclosures and in corporate decision-making. Regulatory initiatives should:

- Integrate algorithmic accountability requirements into existing CFS directives (e.g., CSRD, SEC climate rules).
- Support the development of standardized audit protocols for ESG-focused ML models.
- Encourage sandbox environments for testing

governance frameworks in high-impact sectors.

For Investors and Asset Managers

Due diligence must extend beyond ESG scores to assess algorithmic accountability and the quality of the data underpinning ML-driven CFS claims. Investors should:

- Demand transparency on ML model explainability and bias mitigation in investee companies.
- Develop evaluation frameworks that penalize "automated greenwashing" and reward substantive, verifiable CFS integration.
- Support shareholder resolutions calling for ethical ML governance disclosures.

For Technology Developers and ESG Rating Agencies

Providers of ML-driven CFS tools must prioritize transparency and fairness over secrecy. This includes:

- Publishing model cards detailing data sources, assumptions, and known limitations.
- Engaging in multi-stakeholder co-design processes to mitigate bias and align with diverse CFS perspectives.
- Differentiating clearly between aspirational and performance-based data in scoring methodologies.

Limitations and Future Research Trajectories

While this review synthesizes a growing body of literature, it also highlights several critical limitations in the current research landscape

Methodological and Empirical Gaps

Causal Evidence Scarcity

Most studies show correlation rather than causation between ML adoption and CFS outcomes. There is an urgent need for longitudinal, quasi-experimental studies that isolate ML's impact on long-term financial and ESG performance.

Geographical and Firm-Size Bias

The literature remains dominated by studies from developed economies and large firms, limiting understanding of dynamics in emerging markets and SMEs, where resource constraints and institutional voids pose unique challenges.

Theoretical Development Needs

Future research should develop novel theoretical hybrids that integrate insights from ML ethics, science and technology studies (STS), and sustainable finance. Promising directions include:

Algorithmic Legitimacy

Examining how ML systems gain or lose legitimacy among diverse stakeholders.

Digital Stakeholder Sovereignty

Exploring frameworks for meaningful stakeholder participation in ML governance.

Resilience-Based ML Governance

Developing models that prioritize adaptive capacity over predictive precision.

Emerging Topics Requiring Attention

Generative ML and CFS Reporting

Investigating how large language models may transform CFS disclosures and stakeholder communication.

Blockchain for Algorithmic Auditing

Exploring distributed ledger technologies for creating immutable, auditable trails of ML training data and decision processes.

Participatory ML in Just Transitions

Examining how inclusive design processes can ensure ML supports equitable CFS transitions.

Concluding Synthesis: Beyond the Efficiency–Governance Divide

The central argument from this discussion is that adopting ML alone is insufficient and potentially hazardous without intentional, context-sensitive governance. The proposed four-pillar framework, encompassing transparency, bias auditing, integrated reporting, and stakeholder inclusion, offers a pathway to reconcile the efficiency–governance divide. However, its implementation must be tempered by recognition of contextual factors, including regulatory environments, industry materiality, and organizational culture.

Ultimately, the path to sustainable corporate futures will be increasingly digital. The critical question is whether ML will be harnessed as a tool for authentic, equitable, and resilient value creation or become a new vector for obfuscation, inequality, and short-termism. The answer will depend not on the algorithms themselves but on the wisdom, ethics, and governance of the humans and institutions that deploy them. This review underscores that achieving this alignment requires concerted action across academia, industry, policy, and civil society, as well as collective stewardship of technology in service of sustainable human and planetary flourishing.

Proposed Framework

To reconcile the persistent tension between instrumental efficiency and normative governance, and to guide organizations toward the ethical and effective deployment of ML, we propose an integrative Responsible ML Governance Framework for CFS. This framework is built on four interdependent governance pillars, the effectiveness of which is moderated by key contextual factors (see Figure 1 and Table 2). It moves beyond technical compliance toward a holistic, stakeholder-sensitive model of ML stewardship that aligns algorithmic innovation with the long-term goals of CFS and ESG integrity.

Pillar 1: Transparent and Explainable ML Models

Transparency is not merely a technical feature but a foundational governance requirement. This pillar mandates that ML systems used in CFS contexts be interpretable, auditable, and accountable. Organizations should:

• Prioritize Explainable ML (XML) techniques (e.g., SHAP, LIME) for high-stakes CFS decisions, especially in ESG scoring and risk prediction.

• Maintain comprehensive model documentation (e.g., model cards, data statements) that records data provenance, assumptions, limitations, and intended use cases.

• Establish protocols for third-party algorithmic audits to enable independent verification of model fairness, accuracy, and robustness.

These measures ensure that stakeholders, from boards to regulators to civil society, can understand, question, and ultimately trust ML-driven CFS insights.

Pillar 2: Bias-Audit and Ethical ML Governance

Moving beyond ad-hoc checks, this pillar institutionalizes continuous bias auditing and ethical oversight throughout the ML lifecycle. Key mechanisms include:

• Establishing multidisciplinary ML ethics boards with representation from sustainability, legal, compliance, and impacted communities.

• Conducting regular disparate impact analyses to assess how model outputs affect different stakeholder groups, sectors, and regions.

• Ensuring diversity in ML development teams to mitigate design blind spots and embed pluralistic values into algorithmic systems.

Ethical governance thus transforms ML from a technical project into a corporate asset with clear lines of accountability and redress.

Pillar 3: Integrated ESG-ML Reporting Standards

Rather than creating parallel or opaque ML-driven metrics, this pillar emphasizes alignment with global CFS reporting standards. Organizations should:

• Use ML to enhance, not replace, compliance with frameworks such as the IFRS Sustainability Disclosure Standards, CSRD, and TCFD.

• Disclose the role of ML in data collection, analysis, and reporting within mainstream CFS reports.

• Seek external assurance for ML-generated metrics to ensure they meet the same rigor as traditional financial disclosures.

Integration ensures that ML contributes to a single, coherent, and auditable narrative of corporate performance, reducing the risk of "automated greenwashing" or disclosure fragmentation.

Pillar 4: Stakeholder-Inclusive ML Deployment

Recognizing that ML systems affect and are affected by a wide range of stakeholders, this pillar mandates participatory and inclusive governance. Practical approaches include:

• Convening stakeholder panels to provide input on ML system design, especially for community-impact assessments and social governance tools.

• Involving employee representatives in the oversight of workplace ML applications related to safety, productivity, and well-being.

• Engaging investors and policymakers in transparent dialogue about the capabilities, limitations, and ethical boundaries of sustainability ML.

Inclusion not only enhances legitimacy but also helps identify unintended consequences early, promoting socially robust and ethically aligned systems. Aduloju *et al.* (2025) showed that successful technology integration relies not only on operational efficiency but also on strategic alignment with organizational goals and stakeholder acceptance, underscoring the importance of context and governance in technology-driven business innovation.

Moderating Variables: Contextual Factors

The efficacy of the four pillars is not universal; powerful contextual moderators shape it.

Regulatory Stringency

Strong legal frameworks (e.g., EU AI Act, CSRD) compel adherence to Pillars 1 and 3, while weaker regimes may necessitate voluntary leadership.

Industry ESG Materiality

High-impact sectors (e.g., energy, mining) face greater pressure to implement robust governance (Pillars 2 & 4) due to salient CFS risks.

Data and Technological Infrastructure

Firms with mature data governance and digital literacy can operationalize the pillars more effectively.

Organizational Culture and Leadership

A culture of ethical innovation, supported by board-level ML literacy and commitment, is a critical enabler of responsible deployment.

Visual and Tabular Summary of the Framework

To reconcile the efficiency-governance tension and guide practitioners, we propose an integrative framework (see Figure 1 and Table 2). This framework posits that CFS efficiency is a function of four interdependent governance pillars, with the effectiveness of these pillars contingent on key moderating contextual factors.

Research Gap and Future Agenda

This critical review not only synthesizes existing knowledge but also highlights significant gaps that must be addressed to advance scholarship and practice at the

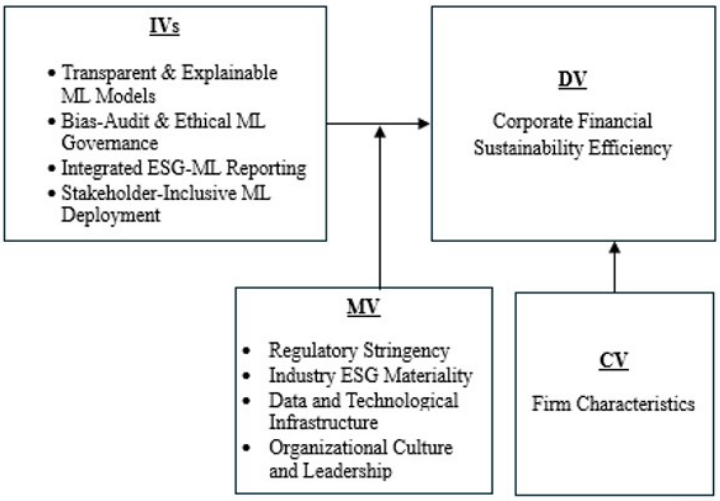

Figure 1: The Responsible ML Governance Framework for CFS

Table 2: Operationalizing the Framework – Key Variables and Measures

Variable Type	Variable	Definition	Exemplary Measures from Literature
Dependent (DV)	Corporate Financial Sustainable Efficiency	Achievement of long-term, resilient financial outcomes integrated with ESG integrity.	LIVA (Wibbens & Siggelkow, 2020), ESG-adjusted ROA (Sandberg *et al.*, 2023), Resilience Index.
Independent (IV1)	Transparent & Explainable ML Models	The interpretability, documentation, and auditability of sustainability-focused ML systems.	Use of SHAP/LIME (Schneider & Schwab, 2025), model card completeness, and audit frequency.
Independent (IV2)	Bias-Audit & Ethical ML Governance	Structured processes to identify, assess, and mitigate algorithmic bias and ethical risks.	Bias impact assessment frequency (Blue *et al.*, 2025), and the presence of an ML ethics board.
Independent (IV3)	Integrated ESG-ML Reporting	Alignment of ML-driven sustainability data with formal reporting standards and assurance.	CSRD/TCFD alignment score (Hąbek, 2025), assurance level of ML-generated metrics.
Independent (IV4)	Stakeholder-Inclusive ML Deployment	Involvement of affected stakeholders in the design, monitoring, and governance of ML systems.	Stakeholder co-design workshops (Seele, 2017), and transparency of ML use cases.
Moderator (MV)	Contextual Factors: • Regulatory Stringency •Indust ry ESG Materiality • Data and Technological Infrastructure •Organizational Culture and Leadership	External and internal conditions that strengthen or weaken the IV-DV relationship.	Regulatory index, industry materiality score (Knoepfel, 2001), and data infrastructure maturity.
Control (CV)	Firm Characteristics	Attributes that may influence the DV and should be accounted for empirically.	Firm size, leverage, ROA, board independence, and CEO duality (Vintilă *et al.*, 2025).

intersection of ML and CFS. Future research should prioritize the following interconnected domains:

Theoretical Development

The field requires novel theoretical hybrids that transcend disciplinary silos. Promising directions include:

Algorithmic Legitimacy
How do ML systems gain, maintain, or lose legitimacy among diverse stakeholders in CFS contexts?

Digital Stakeholder Sovereignty
How can governance frameworks ensure meaningful stakeholder agency in ML design and oversight?

Resilience-Based AI Governance
How might ML systems be designed to prioritize adaptive capacity and systemic resilience over narrow predictive accuracy?

Methodological Innovation
Empirical Rigor Must be Elevated Through Longitudinal and Causal Designs
Quasi-experimental and panel studies are needed to isolate the causal impact of ML governance practices (e.g., the four pillars proposed) on long-term CFS and ESG outcomes.

Cross-National Comparative Research
Studies comparing regulatory, cultural, and market contexts, especially between developed and emerging economies, are essential to understand how institutional environments shape ML adoption and its consequences.

Mixed-Methods Approaches
Qualitative and participatory research can uncover stakeholder perceptions, ethical dilemmas, and contextual nuances that quantitative models may overlook.

Empirical and Practical Priorities
To ensure relevance and inclusivity, research must expand beyond current biases:

ML in SMEs and Emerging Markets
Investigations into how resource-constrained firms navigate ML adoption, and how institutional voids create both risks and opportunities for leapfrogging.

Standardized Toolkits for Auditing and Explainability
The development, validation, and dissemination of practical, open-source tools for bias auditing and XML tailored to ESG and CFS models.

Impact of Generative ML
Research into how large language models and generative ML are reshaping CFS reporting, stakeholder communication, and the risk of synthetic disclosure.

Emerging and Interdisciplinary Frontiers
Blockchain for Algorithmic Accountability
Exploring distributed ledger technologies as mechanisms for creating immutable, auditable records of ML training data and decision pathways.

Participatory ML for Just Transitions
Examining how inclusive, co-design processes can ensure that ML supports equitable CFS transitions, particularly in communities vulnerable to climate and technological disruption.

Psychosocial and Behavioral Dimensions
Examining how ML-driven CFS interfaces influence managerial decision-making, investor behavior, and consumer trust.
Addressing these gaps will require collaborative, interdisciplinary, and transnational research that bridges technical, social, and ethical domains.

CONCLUSION

This review examines ML's dual role in CFS. While ML enhances predictive accuracy and operational efficiency, it also introduces risks of algorithmic bias, opacity, and eroded accountability. Without deliberate governance, ML may undermine CFS by automating inequity and prioritizing short-term efficiency over long-term resilience. We propose a four-pillared governance structure, namely Transparent Models, Bias Auditing, Integrated ESG-ML Reporting, and Stakeholder-Inclusive Deployment, to ensure ML aligns with CFS goals. This strategy calls for corporate governance to modernize, overseeing not only financial resources but also data and algorithms. Its implementation should be flexible and sensitive to varying contexts, guided by regulatory, technological, and cultural influences. CFS is becoming irreversibly digital. The critical question is no longer whether to adopt ML but how to govern it responsibly. The outcome depends on human wisdom and institutional governance, not the technology alone. We call for interdisciplinary research, corporate accountability, informed policymaking, and civil society engagement to ensure ML serves as an architect of sustainable, equitable value creation.

REFERENCES

Adams, C. A., & Whelan, G. (2009). Conceptualising future change in corporate sustainability reporting. *Accounting, Auditing & Accountability Journal, 22*(1), 118–143. https://doi.org/10.1108/09513570910923033

Aduloju, O. D., Adedotun, A. K., & Taiwo, A. A. (2025). An exploratory study on the use of robotics to enhance marketing strategies in business organizations: A case study of selected firms. *American Journal of Data Science and Artificial Intelligence, 1*(1), 27–35. https://doi.org/10.54536/ajdsai.v1i1.4933

Alipasa, C. D. L. (2026). The parents' perception on the extent of implementation of virtual learning: Basis for a faculty e learning training program. *Perspectives in Sustainable Development Studies, 1*(1), 1–11. https://doi.org/10.65401/642700

Allen, F., Qian, J., Shan, C., & Zhu, J. L. (2024). Dissecting the long-term performance of the Chinese stock market. *The Journal of Finance, 79*(2), 993–1054. https://doi.org/10.1111/jofi.13312

Atif, M., Nadarajah, S., & Richardson, G. (2023). Staggered

adoption of stakeholder constituency statutes and corporate cash holdings in the U.S. *Economic Modelling, 124,* 106325. https://doi.org/10.1016/j.econmod.2023.106325

Bernardini, E., Fanari, M., Foscolo, E., & Ruggiero, F. (2024). Environmental data and scores: Lost in translation. *Corporate Social Responsibility and Environmental Management, 31*(5), 4796–4818. https://doi.org/10.1002/csr.2829

Blue, G., Chahrdahcheriki, M., Rezaee, Z., & Khotanlou, M. (2025). A model for predicting creative accounting in emerging economies. *International Journal of Accounting and Information Management, 33*(1), 1–31. https://doi.org/10.1108/IJAIM-09-2023-0240

Cheng, L. T. W., Cheong, T. S., Wojewodzki, M., & Chui, D. K. H. (2025). The effect of ESG divergence on the financial performance of Hong Kong-listed firms: An artificial neural network approach. *Research in International Business and Finance, 73,* 102616. https://doi.org/10.1016/j.ribaf.2024.102616

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147–160. https://doi.org/10.2307/2095101

Downing, N. J. (2025). Missing value imputation in environmental, social, and governance data: An impact on emissions scores. *Finance Research Letters, 85,* 107818. https://doi.org/10.1016/j.frl.2025.107818

Fang, J., Li, J., & Bi, P. (2025). Green paradox of AI: Short-term pain and long-term redemption—The two faces of Chinese enterprises' sustainable development. *Finance Research Letters, 86,* 108544. https://doi.org/10.1016/j.frl.2025.108544

Ferro, A., Marazzina, D., & Stocco, D. (2025). Uncovering ESG ratings: The (im)balance of aspirational and performance features. *Corporate Social Responsibility and Environmental Management, 32*(5), 5895–5917. https://doi.org/10.1002/csr.70007

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

Gribnau, H. (2025). Sustainable tax governance: A shared responsibility. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.5065836

Hąbek, P. (2025). Evaluating ESG software solutions for sustainability reporting in the manufacturing sector. *Management Systems in Production Engineering, 33*(3), 420–432. https://doi.org/10.2478/mspe-2025-0041

Hsieh, M. Y. (2024). An empirical investigation into the enhancement of decision-making capabilities in corporate sustainability leadership through Internet of Things (IoT) integration. *Internet of Things, 28,* 101382. https://doi.org/10.1016/j.iot.2024.101382

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

Khan, S., & Gupta, S. (2025). Boosting the efficacy of green accounting for better firm performance: Artificial intelligence and accounting quality as moderators. *Meditari Accountancy Research, 33*(2), 472–496. https://doi.org/10.1108/MEDAR-02-2024-2379

Knoepfel, I. (2001). Dow Jones Sustainability Group Index: A global benchmark for corporate sustainability. *Corporate Environmental Strategy, 8*(1), 6–15. https://doi.org/10.1016/s1066-7938(00)00089-0

Li, J., Wu, T., Hu, B., Pan, D., & Zhou, Y. (2025). Artificial intelligence and corporate ESG performance. *International Review of Financial Analysis, 102,* 104036. https://doi.org/10.1016/j.irfa.2025.104036

Lin, B. W., Chen, C. J., & Wu, H. L. (2006). Patent portfolio diversity, technology strategy, and firm value. *IEEE Transactions on Engineering Management, 53*(1), 17–26. https://doi.org/10.1109/tem.2005.861813

Martin, G., Farndale, E., Paauwe, J., & Stiles, P. G. (2016). Corporate governance and strategic human resource management: Four archetypes and proposals for a new approach to corporate sustainability. *European Management Journal, 34*(1), 22–35. https://doi.org/10.1016/j.emj.2016.01.002

Naveed, K., Farooq, M. B., Zahir-ul-Hassan, M. K., & Rauf, F. (2025). AI adoption, ESG disclosure quality and sustainability committee heterogeneity: Evidence from Chinese companies. *Meditari Accountancy Research, 33*(2), 708–732. https://doi.org/10.1108/MEDAR-02-2024-2374

Sandberg, H., Alnoor, A., & Tiberius, V. (2023). Environmental, social, and governance ratings and financial performance: Evidence from the European food industry. *Business Strategy and the Environment, 32*(4), 2471–2489. https://doi.org/10.1002/bse.3259

Schneider, J. C., & Schwab, B. (2025). Advancing loss reserving: A hybrid neural network approach for individual claim development prediction. *Journal of Risk and Insurance, 92*(2), 389–423. https://doi.org/10.1111/jori.12501

Seele, P. (2017). Predictive sustainability control: A review assessing the potential to transfer big data driven 'predictive policing' to corporate sustainability management. *Journal of Cleaner Production, 153,* 673–686. https://doi.org/10.1016/j.jclepro.2016.10.175

Shabbir, M. S. (2025). Corporate sustainability reimagined: A bibliometric–systematic literature review of governance, technology, and stakeholder-driven strategies for SDG impact. *Business Strategy and the Environment, 34*(7), 9203–9222. https://doi.org/10.1002/bse.70070

Shilbayeh, S. A., & Grassa, R. (2024). Creditworthiness pattern prediction and detection for GCC Islamic banks using machine learning techniques. *International Journal of Islamic and Middle Eastern Finance and Management, 17*(2), 345–365. https://doi.org/10.1108/IMEFM-02-2023-0057

Spagnuolo, F., Casciello, R., Martino, I., & Meucci, F. (2025). Exploring the impact of artificial intelligence on the pursuit of SDGs: Evidence from European state-owned enterprises. *Corporate Social Responsibility*

and Environmental Management, 32(2), 1987–2001. https://doi.org/10.1002/csr.3047

Strätling, R. (2007). The legitimacy of corporate social responsibility. *Corporate Ownership and Control, 4*(4), 80–88. https://doi.org/10.22495/cocv4i4p6

Suárez Giri, F., & Sánchez-Chaparro, T. (2024). Unveiling the blackbox within ESG ratings' blackbox: Toward a framework for analyzing AI adoption and its impacts. *Business Strategy and Development, 7*(4), e70038. https://doi.org/10.1002/bsd2.70038

Subramaniam, R. K., Samuel, S. D., Seera, M., & Alam, N. (2024). Utilising machine learning for corporate social responsibility (CSR) and environmental, social, and governance (ESG) evaluation: Transitioning from committees to climate. *Sustainable Futures, 8,* 100329. https://doi.org/10.1016/j.sftr.2024.100329

Sulimany, H. G., Ramakrishnan, S., Chaudhry, A., & Bazhair, A. H. (2021). Impact of corporate governance and financial sustainability on shareholder value. *Estudios de Economía Aplicada, 39*(4). https://doi.org/10.25115/eea.v39i4.4318.

Vintilă, G., Onofrei, M., & Vintilă, A. I. (2025). Boosting IT companies' performance through corporate governance standards: An empirical analysis. *Oeconomia Copernicana, 16*(2), 761–809. https://doi.org/10.24136/oc.3627

Wibbens, P. D., & Siggelkow, N. (2020). Introducing LIVA to measure long-term firm performance. *Strategic Management Journal, 41*(5), 867–890. https://doi.org/10.1002/smj.3114

Zhai, S., & Zhang, Z. (2023). Read the news, not the books: Forecasting firms' long-term financial performance via deep text mining. *ACM Transactions on Management Information Systems, 14*(1). https://doi.org/10.1145/3533018

Zhao, J., & Gómez Fariñas, B. (2023). Artificial intelligence and sustainable decisions. *European Business Organization Law Review, 24*(1), 1–39. https://doi.org/10.1007/s40804-022-00262-2

Zhou, Z., Zhou, X., Zhang, X., & He, Q. (2025). Not all sparks ignite the same flame: Firm AI innovation and ESG performance. *Journal of Business Research, 201,* 115738. https://doi.org/10.1016/j.jbusres.2025.115738