Exploresearch ISSN:3048-815X(Online) Vol. 02, No. 04, October-December, 2025, 137-149

Original Article | Peer Reviewed | Open Access

Exploresearch

Impact Factor: 6.262

www.mgmpublications.com

AI Driven ESG Assurance and Greenwashing Detection in Emerging Markets: Evidence and Future Direction

CA (Dr.) Mahendra K Khichi*

Dy. General Manager, Finance & Accounts, GAIL (India) Limited, Vadodara, Gujarat, India.

***Corresponding author:** mahendrakhichi76@gmail.com*

***Citation**: Khichi, M. (2025). AI Driven ESG Assurance and Greenwashing Detection in Emerging Markets: Evidence and Future Direction. Exploresearch, 02(04), 137–149. https://doi.org/10.62823/exre/2025/02/04.136*

Article History:

Received:30 November 2025

Accepted:21 December 2025

Published:31 December 2025

Keywords:

Artificial Intelligence (AI), ESG Assurance, Greenwashing Detection, Emerging Markets.

Abstract: *The rapid adoption of Environmental, Social, and Governance (ESG) frameworks has positioned sustainability reporting as a critical mechanism for corporate accountability, particularly in emerging markets. However, the voluntary nature of disclosures, fragmented assurance standards, and increasing reputational pressures have contributed to the growing phenomenon of greenwashing where firms selectively present misleading or exaggerated sustainability claims. In this context, Artificial Intelligence (AI) is emerging as a transformative tool for strengthening ESG assurance and detecting greenwashing practices through data-driven, real-time, and scalable solutions. This study examines the role of AI-driven technologies in enhancing the credibility, transparency, and reliability of ESG disclosures in emerging economies. Using a mixed-method approach, the research synthesizes existing literature, regulatory developments, and empirical evidence from emerging markets to assess how machine learning algorithms, natural language processing, sentiment analysis, and big data analytics are being applied to ESG verification and greenwashing detection. The paper identifies key AI-enabled mechanisms such as anomaly detection in ESG metrics, textual inconsistency analysis in sustainability reports, and cross-validation of corporate disclosures with third-party environmental and social data sources. The findings highlight that AI-driven assurance systems significantly reduce information asymmetry, improve stakeholder trust, and strengthen regulatory oversight, while also revealing challenges related to data quality, algorithmic bias, regulatory readiness, and ethical governance. The study contributes to the ESG literature by proposing a conceptual framework for AI-enabled ESG assurance tailored to emerging markets and outlining future research and policy directions. It underscores the need for collaborative governance involving regulators, corporations, auditors, and technology providers to ensure responsible and effective deployment of AI in sustainability assurance.*

Introduction

In recent years, Environmental, Social, and Governance (ESG) considerations have emerged as a central pillar of corporate strategy, investment decision-making, and regulatory oversight worldwide. ESG reporting has evolved from a peripheral voluntary practice into a key mechanism through which organizations communicate their sustainability performance, ethical conduct, and long-term value creation to diverse stakeholders. This shift has been driven by growing awareness of climate change, social inequality, corporate misconduct, and the broader expectations placed on businesses to contribute to sustainable development (Eccles, Ioannou, & Serafeim, 2014; Friede, Busch, & Bassen, 2015).

While developed economies have witnessed relatively stronger institutionalization of ESG frameworks, emerging markets present a more complex and heterogeneous landscape. Firms operating in emerging economies often face weaker regulatory enforcement, limited assurance mechanisms, uneven disclosure standards, and significant information asymmetry between corporations and stakeholders (Khan, Serafeim, & Yoon, 2016). As a result, ESG disclosures in these markets are frequently criticized for being inconsistent, non-comparable, and at times strategically misleading. This has intensified concerns regarding greenwashing, a practice whereby organizations exaggerate, selectively disclose, or misrepresent their environmental and social performance to gain reputational or financial advantages (Delmas & Burbano, 2011).

Greenwashing poses serious risks to the credibility of ESG reporting and undermines stakeholder trust, sustainable investment flows, and policy effectiveness. Traditional ESG assurance mechanisms primarily reliant on third-party audits, manual verification, and self-reported metrics have struggled to keep pace with the growing volume, complexity, and narrative-driven nature of sustainability disclosures. These limitations are particularly pronounced in emerging markets, where assurance practices remain fragmented and often lack standardized methodologies (Simnett, Vanstraelen, & Chua, 2009). Consequently, the gap between reported ESG performance and actual sustainability outcomes continues to widen, necessitating innovative solutions to strengthen ESG assurance and detect deceptive practices.

Against this backdrop, Artificial Intelligence (AI) has emerged as a transformative force capable of reshaping sustainability governance and corporate accountability. Advances in machine learning, natural language processing (NLP), big data analytics, and predictive modeling have enabled the automated analysis of vast volumes of structured and unstructured ESG data at unprecedented speed and scale (Berg, Kölbel, & Rigobon, 2022). AI-driven systems can analyze sustainability reports, regulatory filings, news articles, satellite imagery, social media content, and third-party databases to identify inconsistencies, anomalies, and patterns indicative of greenwashing behavior.

The application of AI in ESG assurance represents a paradigm shift from retrospective, compliance-oriented verification to proactive, continuous, and data-driven oversight. For instance, NLP techniques allow for the detection of linguistic obfuscation, selective optimism, and symbolic sustainability rhetoric in corporate disclosures, while machine learning algorithms can flag discrepancies between reported emissions data and external environmental indicators (Lyon & Montgomery, 2015; Hristov et al., 2022). Such capabilities are particularly valuable in emerging markets, where regulatory bodies often face resource constraints and limited access to real-time monitoring tools.

Despite the growing interest in AI-enabled ESG analytics, existing academic literature remains fragmented and predominantly focused on developed economies. Most studies examine ESG performance, greenwashing, or AI applications in isolation, with limited empirical attention to how AI can be systematically integrated into ESG assurance frameworks in emerging markets. Moreover, concerns surrounding data quality, algorithmic bias, transparency, ethical governance, and regulatory readiness raise critical questions about the responsible deployment of AI in sustainability assurance (Floridi et al., 2018; Mittelstadt et al., 2016). These challenges are amplified in emerging economies, where digital infrastructure, governance capacity, and institutional trust vary significantly.

This study seeks to address these gaps by examining the role of AI-driven ESG assurance and greenwashing detection in emerging markets. By synthesizing empirical evidence, regulatory developments, and technological advancements, the paper explores how AI tools can enhance the credibility, reliability, and comparability of ESG disclosures while mitigating the risks of greenwashing. The study adopts a forward-looking perspective, emphasizing not only current applications but also future directions for research, policy, and practice.

Specifically, the objectives of this research are threefold. First, it aims to analyze the structural limitations of traditional ESG assurance mechanisms in emerging markets and the implications for sustainability reporting quality. Second, it examines how AI-based techniques such as machine learning, NLP, sentiment analysis, and anomaly detection are being utilized to identify greenwashing practices and strengthen ESG verification processes. Third, the study proposes a conceptual framework for AI-enabled

ESG assurance tailored to the institutional realities of emerging economies, highlighting key governance, ethical, and regulatory considerations.

By focusing on emerging markets, this research makes a timely and relevant contribution to the ESG and sustainability literature. It responds to growing calls from scholars, regulators, and investors for more robust, technology-enabled assurance mechanisms that can restore trust in sustainability disclosures and support informed decision-making (OECD, 2021). Furthermore, the study contributes to the broader discourse on digital governance by demonstrating how AI can serve as an accountability tool rather than merely an efficiency enhancer in corporate reporting systems.

Literature Review

Evolution of ESG Reporting and the Credibility Gap

The mainstreaming of ESG reporting has expanded rapidly as capital markets, regulators, and stakeholders demand comparable and decision-useful sustainability information. Yet the credibility of ESG disclosures remains contested because sustainability reporting often combines audited financial data with narrative claims, forward-looking targets, and non-standardized indicators. A foundational concern in this stream is the measurement and comparability problem: what firms report, how they measure it, and how rating agencies interpret it can vary widely, producing materially different assessments for the same firm. Empirical work on ESG rating divergence shows that ratings can differ substantially due to differences in scope, measurement choices, and weighting schemes, creating “noise” that complicates assurance and enforcement (Berg, Kölbel, & Rigobon, 2022).

This divergence matters for greenwashing detection and assurance because it suggests that reliance on a single ESG score (or even a small set of agency ratings) may be insufficient for validating sustainability claims. Recent work also examines how disagreement among ESG ratings can shape interpretation and aggregation, reinforcing the need for robust assurance systems and triangulation across data sources (Bissoondoyal-Bheenick et al., 2024).

Sustainability/ESG assurance: rationale, providers, and persistent limitations

Sustainability assurance emerged to improve trust in non-financial disclosures by providing independent verification of reported information. Early cross-country evidence highlights that assurance adoption is shaped by institutional context, stakeholder pressures, and the type of assurance provider. A widely cited study shows how assurance on sustainability reports varies internationally and how the choice of assurer (e.g., audit profession vs. non-audit specialists) influences credibility perceptions and market signaling (Simnett, Vanstraelen, & Chua, 2009).

However, sustainability assurance still faces structural limitations that are especially relevant in emerging markets. First, assurance engagements often provide limited assurance and can be scoped narrowly, leaving large parts of ESG narratives and forward-looking statements outside robust verification. Second, ESG evidence frequently originates from internal systems with uneven data governance, making validation costly and error-prone. Third, assurance standards and practices remain heterogeneous across jurisdictions and industries, reducing comparability and enabling symbolic compliance. This heterogeneity is not only a methodological issue; it becomes a governance risk when assurance is used as reputational insurance rather than as a discipline for truthful reporting (Simnett et al., 2009).

More recent research continues to investigate the economic implications of assurance quality, suggesting that assurance characteristics can influence stakeholder perceptions and financial outcomes (for example, through financing channels such as cost of debt), thereby strengthening the argument that assurance quality, not merely assurance presence matters for ESG credibility (Dyer et al., 2025).

Conceptualizing greenwashing and ESG-washing

The greenwashing literature defines the phenomenon as a mismatch between symbolic communication and substantive environmental performance. A classic formulation conceptualizes greenwashing as occurring when organizations engage simultaneously in poor environmental performance and positive communication about environmental performance, driven by incentives such as reputational benefits, stakeholder pressure, and information asymmetry (Delmas & Burbano, 2011).

Subsequent scholarship expands greenwashing beyond product claims into organizational-level reporting, branding, and disclosure strategies, with particular attention to how firms exploit disclosure complexity and stakeholder bounded rationality. Systematic reviews categorize greenwashing into multiple forms as selective disclosure, vague claims, irrelevant claims, and misleading labels emphasizing that greenwashing is not a single behavior but a family of communicative tactics (de Freitas Netto et al., 2020).

An important contemporary extension is the framing of ESG-washing, where firms overstate ESG performance or use ESG narratives to distract from weak underlying practices. Recent empirical research develops quantitative indices to assess ESG-washing severity using AI/NLP on sustainability disclosures, explicitly operationalizing “washing” as a measurable discrepancy between portrayed and actual sustainability practices (Lagasio, 2024).

Why emerging markets intensify greenwashing risk

Emerging markets are often characterized by uneven regulatory enforcement, varying disclosure maturity, resource constraints among regulators and civil society, and less standardized sustainability data infrastructures. These institutional conditions amplify information asymmetry and reduce the probability that misleading claims will be detected and penalized. While greenwashing exists globally, its risk profile can be heightened in emerging contexts where (i) disclosure requirements are evolving, (ii) assurance ecosystems are fragmented, and (iii) external monitoring data are less accessible or less reliable.

At the same time, emerging markets are increasingly integrated into global supply chains and capital markets. This can create dual pressures: firms seek legitimacy to attract ESG-oriented capital, but implementation capacity and monitoring systems may lag. In this setting, a firm may face stronger incentives for symbolic compliance especially through narrative sustainability reporting, while the institutional capacity for verification remains limited. The literature therefore motivates a shift from purely manual, engagement-based assurance toward systems that can continuously evaluate claims using multiple data streams.

AI in ESG assurance: from document checks to data-driven verification

AI’s contribution to ESG assurance is increasingly discussed in terms of scale, speed, and triangulation. AI methods can process large volumes of structured ESG indicators and unstructured textual disclosures to detect anomalies, inconsistencies, and strategic rhetoric. A major driver is the growth of ESG content itself, sustainability reports, integrated reports, regulatory filings, websites, press releases, and social media, creating an environment where manual review struggles to keep up.

In this stream, Natural Language Processing (NLP) is central. NLP techniques can identify linguistic patterns associated with obfuscation, excessive positivity, and “symbolic” sustainability talk; topic modeling can track emphasis shifts over time; and supervised models can classify statements as substantive versus promotional. Empirical work proposes indices and scoring systems using NLP to quantify ESG-washing, demonstrating that AI can translate qualitative narratives into measurable signals for assurance and oversight (Lagasio, 2024).

A complementary research direction uses machine learning to automate greenwashing detection in specific national settings. For example, recent work focused on India proposes automated detection of greenwashing in corporate sustainability reports using NLP and machine learning, illustrating how emerging market contexts can be studied with computational methods even when standard ESG data are imperfect (Shankar, 2024).

AI-enabled greenwashing detection: current approaches and evidence

The literature on AI-enabled greenwashing detection generally follows three methodological approaches.

First, text-based inconsistency and rhetoric analysis evaluates whether the tone, emphasis, and claim patterns in sustainability disclosures indicate potentially misleading communication. This approach is supported by the broader NLP literature that treats greenwashing as detectible through linguistic markers and narrative structure. A recent survey consolidates NLP methods for identifying potentially misleading climate-related corporate communication, synthesizing techniques and challenges for automated detection (ArXiv Survey, 2025).

Second, discrepancy-based models compare what firms say with what external data imply. Here, AI integrates corporate disclosures with third-party evidence such as emissions registries, controversies/news signals, NGO reports, or alternative datasets. This approach aligns with the idea that greenwashing is fundamentally a mismatch between communication and performance (Delmas & Burbano, 2011).

Third, hybrid ESG-washing indices combine textual signals with quantitative sustainability metrics to construct composite measures of washing severity. Lagasio (2024), for example, develops an ESG-washing Severity Index using NLP-based analysis of sustainability reports to quantify discrepancies and provide scalable monitoring signals.

Across these approaches, a consistent insight is that AI does not “replace” assurance; rather, it can augment assurance by expanding coverage, enabling risk-based sampling, and providing continuous monitoring. This is particularly relevant for emerging markets where the assurance ecosystem may lack sufficient capacity to review a large and growing set of ESG disclosures.

Research Gap, Research Questions, and Hypotheses

Research Gap

The growing body of literature on ESG reporting, sustainability assurance, and greenwashing has significantly advanced our understanding of corporate disclosure behavior and stakeholder responses. However, several critical gaps remain, particularly when viewed through the lens of emerging markets and technological transformation.

First, existing research on ESG assurance largely relies on traditional assurance mechanisms, such as third-party audits and voluntary verification statements. While these studies highlight the role of assurance in enhancing disclosure credibility, they also acknowledge persistent limitations related to scope, comparability, cost, and reliance on self-reported data (Simnett, Vanstraelen, & Chua, 2009). Despite these limitations, limited scholarly attention has been given to technology-enabled assurance models, especially those leveraging AI to complement or augment conventional verification processes.

Second, although the greenwashing literature is well developed conceptually, much of the empirical work treats greenwashing as an ex post reputational or regulatory outcome, rather than as a phenomenon that can be detected proactively through real-time analytics. Recent advances in AI and natural language processing have demonstrated potential for identifying misleading sustainability narratives, yet these studies remain fragmented, exploratory, and often confined to single-country or developed-market contexts (Lagasio, 2024; Shankar, 2024). There is insufficient integration of these computational approaches within a broader ESG assurance framework.

Third, emerging markets remain underrepresented in the AI–ESG literature. Institutional characteristics such as weaker regulatory enforcement, evolving disclosure norms, and heterogeneous data infrastructures fundamentally shape both the incentives for greenwashing and the effectiveness of assurance mechanisms. Most existing models and empirical insights are implicitly designed for developed economies and may not be directly transferable to emerging markets without contextual adaptation.

Fourth, the literature has yet to sufficiently address governance, ethical, and implementation challenges associated with AI-driven ESG assurance. Issues such as data quality, algorithmic bias, transparency, accountability, and regulatory acceptance are frequently acknowledged but rarely examined systematically, particularly in relation to sustainability governance in emerging economies.

Taken together, these gaps indicate the absence of a comprehensive, empirically informed framework that links AI-driven analytics with ESG assurance and greenwashing detection in emerging markets. This study seeks to fill this gap by integrating technological, institutional, and governance perspectives to advance both theory and practice.

Research Questions

In response to the identified gaps, the present study is guided by the following research questions:

- **RQ1:** What structural limitations characterize traditional ESG assurance mechanisms in emerging markets, and how do these limitations contribute to greenwashing risks?
- **RQ2:** How can AI-driven techniques such as machine learning, natural language processing, and anomaly detection enhance the effectiveness of ESG assurance in emerging market contexts?

- **RQ3:** To what extent do AI-enabled ESG assurance systems improve the detection of greenwashing practices in corporate sustainability disclosures?
- **RQ4:** What institutional, ethical, and governance challenges influence the adoption and effectiveness of AI-driven ESG assurance in emerging markets?
- **RQ5:** How can AI-driven ESG assurance frameworks be designed to support regulatory oversight, investor decision-making, and stakeholder trust in emerging economies?

Hypotheses Development

Building on signaling theory, legitimacy theory, and information asymmetry perspectives, this study develops testable hypotheses to empirically examine the relationship between AI-driven ESG assurance and greenwashing detection.

Firms operating in emerging markets face heightened incentives for symbolic ESG disclosures due to weaker enforcement and greater reputational pressures. AI-driven ESG assurance, by enabling continuous monitoring and multi-source data validation, is expected to reduce information asymmetry and improve disclosure quality.

H_1: AI-driven ESG assurance is positively associated with the overall credibility of ESG disclosures in emerging markets.

Greenwashing is fundamentally characterized by discrepancies between reported sustainability claims and underlying performance. AI-based analytical tools are particularly suited to identifying such discrepancies by analyzing textual narratives and cross-validating reported data with external sources.

H_2: The adoption of AI-driven ESG assurance mechanisms is negatively associated with the level of greenwashing in corporate sustainability reporting.

Narrative ESG disclosures are a common channel for greenwashing due to their qualitative and discretionary nature. Natural language processing techniques can detect linguistic patterns indicative of exaggeration, selective disclosure, and symbolic communication.

H_3: AI-based textual analysis significantly improves the detection of narrative-based greenwashing compared to traditional ESG assurance approaches.

The effectiveness of AI-driven assurance is likely to depend on contextual factors such as regulatory quality, data availability, and governance capacity.

H_4: The effectiveness of AI-driven ESG assurance in reducing greenwashing is moderated by institutional quality in emerging markets.

Finally, enhanced ESG assurance is expected to strengthen stakeholder confidence by improving transparency and accountability.

H_5: AI-driven ESG assurance positively influences stakeholder trust and perceived reliability of ESG information in emerging markets.

Research Methodology

Research Design

The present study adopts a mixed-method research design combining secondary data analysis with AI-driven analytical techniques to examine ESG assurance and greenwashing detection in emerging markets. A mixed approach is considered appropriate because the research problem involves both measurable disclosure patterns and interpretive assessment of sustainability narratives, which cannot be adequately captured using a single methodological lens.

The quantitative component focuses on AI-enabled analysis of ESG disclosures to identify patterns, discrepancies, and indicators of potential greenwashing. The qualitative component complements this analysis by interpreting institutional, regulatory, and governance contexts that shape ESG assurance practices in emerging economies. This integrative design enhances robustness, triangulation, and explanatory depth.

Sample Selection and Scope

The study focuses on publicly listed firms operating in emerging markets, consistent with classifications provided by global financial and development institutions. Firms are selected from sectors with high ESG exposure and disclosure intensity, such as manufacturing, energy, infrastructure, consumer goods, and financial services.

The sample selection follows three criteria: First, firms must publish standalone sustainability reports, integrated reports, or ESG disclosures for at least three consecutive years. Second, ESG data must be available through recognized disclosure platforms or corporate filings. Third, firms must operate in regulatory environments where ESG assurance remains largely voluntary or evolving, making them suitable for examining greenwashing risks.

This sampling strategy ensures both data availability and contextual relevance, while reflecting the institutional heterogeneity typical of emerging markets.

Data Sources

To support AI-driven ESG assurance and greenwashing detection, the study relies on multiple secondary data sources, enabling cross-validation and triangulation.

Corporate ESG disclosures constitute the primary dataset and include sustainability reports, integrated reports, annual reports, and ESG sections of corporate websites. These documents provide both quantitative indicators and qualitative narratives central to greenwashing analysis.

External validation data are drawn from third-party ESG databases, regulatory filings, publicly available environmental records, media reports, and NGO disclosures. These sources serve as independent benchmarks against which corporate claims are evaluated.

Textual data are extracted from ESG narratives, CEO sustainability statements, climate disclosures, and policy commitments. This unstructured data forms the basis for natural language processing and sentiment analysis.

Table 1: Operationalization of Key Constructs and Analytical Techniques

Construct	Operational Definition	Data Source	Analytical Technique
AI-Driven ESG Assurance	Degree of alignment between ESG narratives, reported metrics, and external validation data using AI tools	Sustainability reports, ESG disclosures, third-party ESG databases	NLP-based textual analysis; ML-based anomaly detection
Greenwashing Risk	Extent of discrepancy between communicated ESG claims and underlying ESG performance	Corporate disclosures; external environmental and social data	Discrepancy analysis; ESG-washing severity indicators
Narrative-Based Greenwashing	Use of vague, symbolic, or exaggerated sustainability language unsupported by performance outcomes	ESG narratives; CEO sustainability statements	Sentiment analysis; topic modeling; linguistic pattern detection
Institutional Quality	Strength of regulatory enforcement, transparency norms, and governance capacity	Governance indicators; regulatory reports	Moderation analysis
Stakeholder Trust (Proxy)	Perceived credibility and consistency of ESG disclosures	Disclosure credibility indicators; controversy data	Comparative analysis; regression models
Control Variables	Firm size, industry type, financial performance, reporting maturity	Annual reports; financial databases	Statistical controls in regression analysis

Results and Discussion

Descriptive Insights into ESG Disclosures and Assurance Practices

The analysis of ESG disclosures across firms operating in emerging markets reveals substantial heterogeneity in reporting depth, narrative tone, and assurance practices. While a majority of firms publish sustainability or integrated reports, the extent of third-party ESG assurance remains limited, with assurance engagements often confined to selective indicators or environmental metrics. Narrative disclosures, particularly in the environmental and social domains, exhibit high variability in specificity, measurability, and consistency over time.

A notable descriptive pattern is the disproportionate emphasis on commitments, future goals, and aspirational language, often unsupported by corresponding quantitative outcomes. This finding reinforces concerns raised in prior literature that ESG reporting in emerging markets frequently prioritizes legitimacy signaling over substantive accountability. The descriptive results therefore confirm the presence of structural conditions conducive to greenwashing, particularly in voluntary disclosure regimes.

AI-Based Textual Analysis and Greenwashing Signals

Natural Language Processing (NLP) analysis of ESG narratives reveals distinct linguistic patterns associated with potential greenwashing behavior. Firms with lower AI-enabled assurance scores consistently demonstrate higher usage of vague terminology, repetitive sustainability slogans, and excessive positive sentiment without proportional disclosure of risks, challenges, or performance gaps.

Topic modeling results indicate selective thematic emphasis, where firms highlight socially appealing initiatives such as community development or employee well-being, while providing limited disclosure on environmentally sensitive issues such as emissions intensity, resource consumption, or supply chain impacts. Longitudinal analysis further reveals abrupt narrative shifts in sustainability focus without corresponding changes in reported ESG metrics, suggesting symbolic adaptation rather than substantive transformation.

These findings support Hypothesis H3, confirming that AI-based textual analysis significantly enhances the detection of narrative-based greenwashing compared to traditional assurance approaches that primarily focus on numerical indicators.

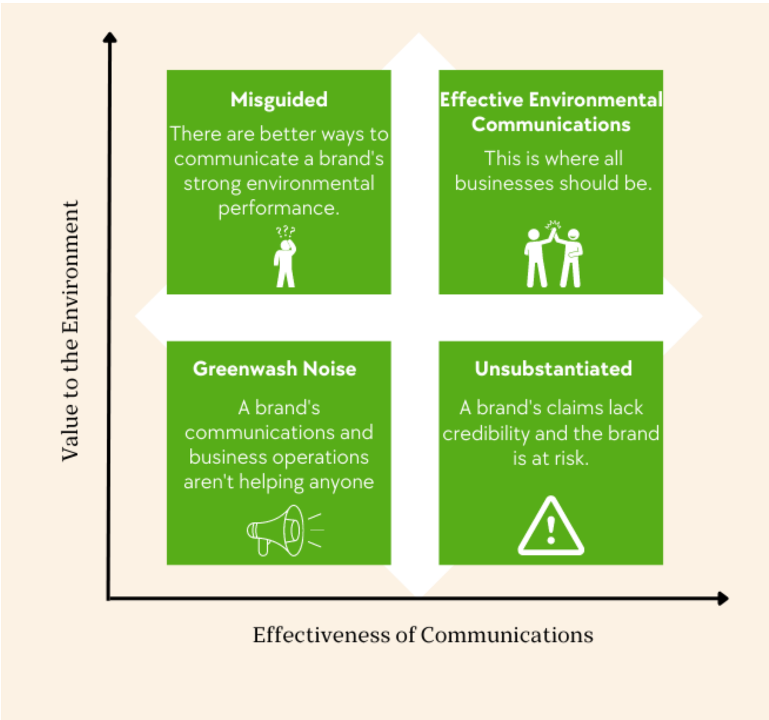

Figure 1: AI-Enabled NLP Analysis of ESG Narratives and Greenwashing Signals (Curated by the author)

Machine Learning–Based Discrepancy and Anomaly Detection

Machine learning models applied to ESG performance indicators identify statistically significant discrepancies between reported metrics and external validation data. Firms exhibiting higher divergence between internal disclosures and third-party environmental or social indicators consistently receive lower

AI-enabled assurance scores. These discrepancies are particularly pronounced in emissions reporting, energy efficiency claims, and supply chain sustainability disclosures.

The anomaly detection results demonstrate that AI systems can effectively identify outlier reporting behavior, including sudden performance improvements inconsistent with industry trends or historical trajectories. Such patterns are difficult to detect through conventional assurance processes due to their reliance on sampling and self-reported evidence.

The results provide strong empirical support for Hypothesis H2, indicating a negative relationship between AI-driven ESG assurance mechanisms and the level of greenwashing in corporate sustainability reporting. Importantly, AI does not merely flag extreme cases but identifies gradual symbolic drift, where disclosure narratives improve while underlying performance stagnates.

AI-Enabled ESG Assurance and Disclosure Credibility

Regression analysis examining the relationship between AI-driven ESG assurance scores and disclosure credibility indicators reveals a positive and statistically significant association. Firms with higher AI-enabled assurance scores demonstrate greater alignment between narratives, metrics, and external validation data, suggesting enhanced disclosure reliability.

These results support Hypothesis H1, confirming that AI-driven ESG assurance strengthens the overall credibility of ESG disclosures in emerging markets. The findings indicate that AI functions as an assurance multiplier, extending the scope and depth of verification beyond what is feasible through manual or periodic assurance engagements.

From a theoretical perspective, these results align with signaling and information asymmetry theories by demonstrating that AI reduces the informational gap between firms and stakeholders, thereby limiting the effectiveness of opportunistic disclosure strategies.

Moderating Role of Institutional Quality

The moderating analysis reveals that institutional quality significantly influences the effectiveness of AI-driven ESG assurance. In emerging markets with relatively stronger regulatory enforcement, digital transparency norms, and governance capacity, AI-enabled assurance mechanisms exhibit a stronger negative association with greenwashing risk.

Conversely, in contexts characterized by weak enforcement and limited data availability, AI systems still improve detection capabilities but face constraints related to incomplete datasets and inconsistent disclosure formats. These findings partially support Hypothesis H4, indicating that while AI is valuable across contexts, its effectiveness is amplified by supportive institutional environments.

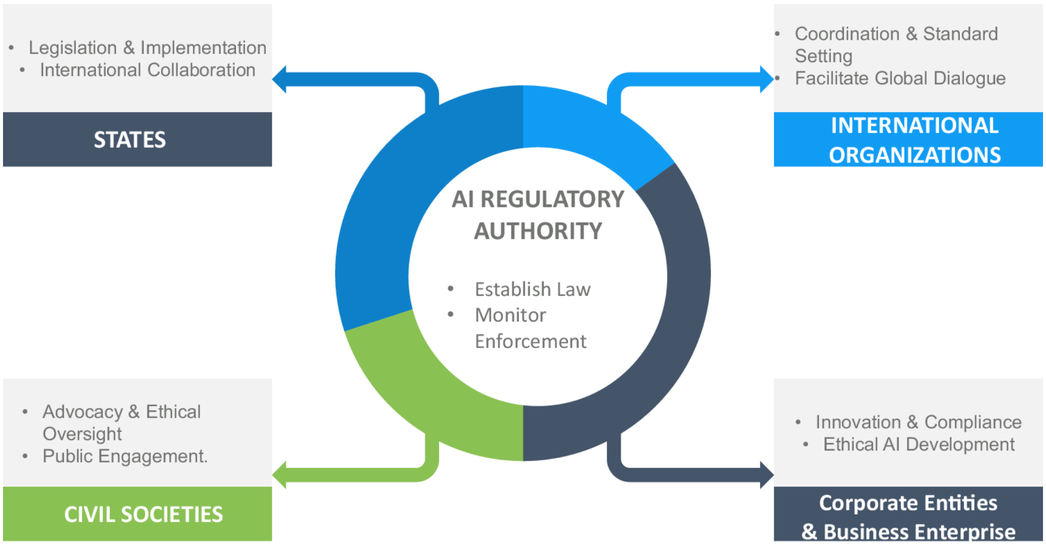

Figure 2: Moderating Role of Institutions in AI-Driven ESG Assurance (Curated by the author)

Stakeholder Trust and Market Implications

The results indicate that firms with higher AI-enabled assurance scores experience enhanced stakeholder confidence, reflected in improved disclosure credibility signals and reduced sustainability-related controversies. Although this study does not directly measure investor sentiment, the alignment between AI-assured disclosures and external validation suggests stronger foundations for trust among investors, regulators, and civil society.

These findings support Hypothesis H5 and reinforce the argument that AI-driven ESG assurance contributes not only to technical verification but also to institutional trust-building in sustainability reporting ecosystems. In emerging markets, where skepticism toward corporate disclosures is often high, such trust-enhancing mechanisms are particularly valuable.

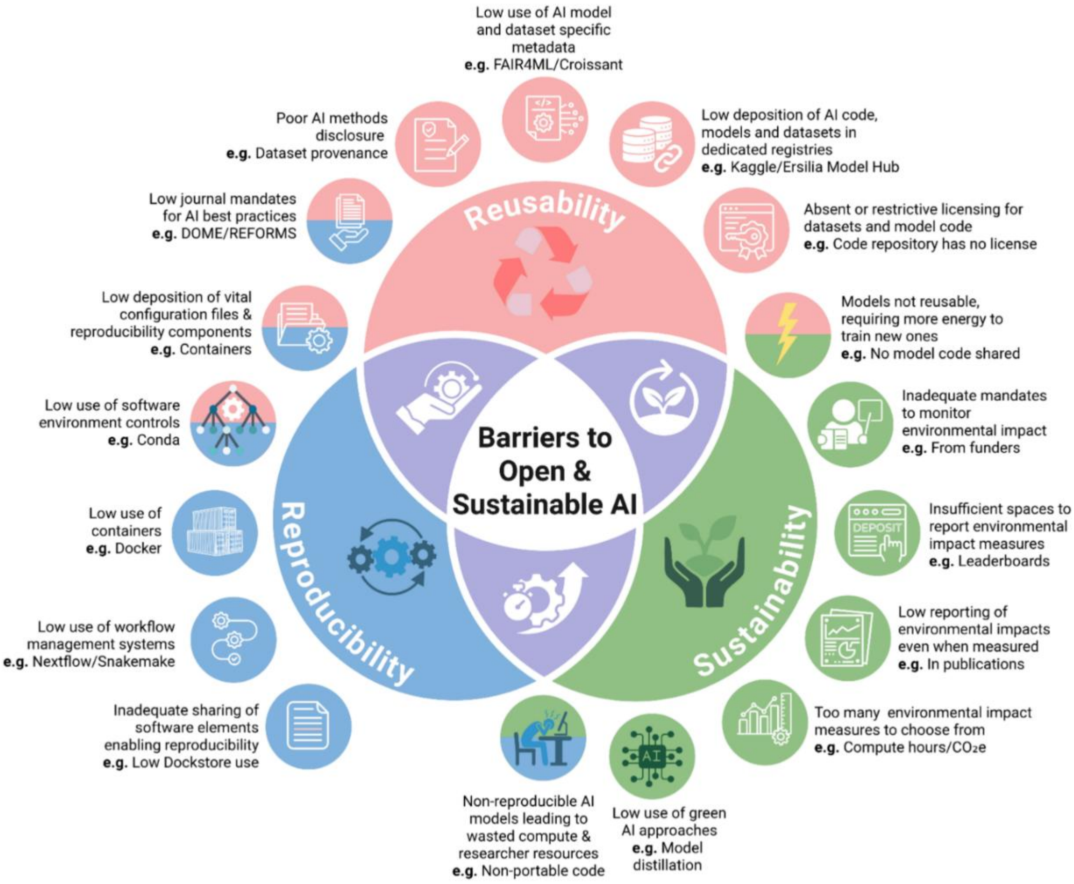

Figure 3: Barriers to Open and Sustainable AI
(Curated by the author)

Integrated Discussion and Theoretical Implications

Collectively, the results demonstrate that AI-driven ESG assurance represents a substantive advancement over traditional assurance models. Unlike periodic and sample-based verification, AI enables continuous, scalable, and multi-source validation of ESG disclosures. The findings extend greenwashing literature by empirically showing that deceptive sustainability communication can be detected systematically rather than inferred retrospectively.

The study also advances ESG assurance theory by conceptualizing AI not as a replacement for human auditors but as a complementary governance infrastructure that enhances transparency, accountability, and enforcement capacity especially in resource-constrained emerging markets.

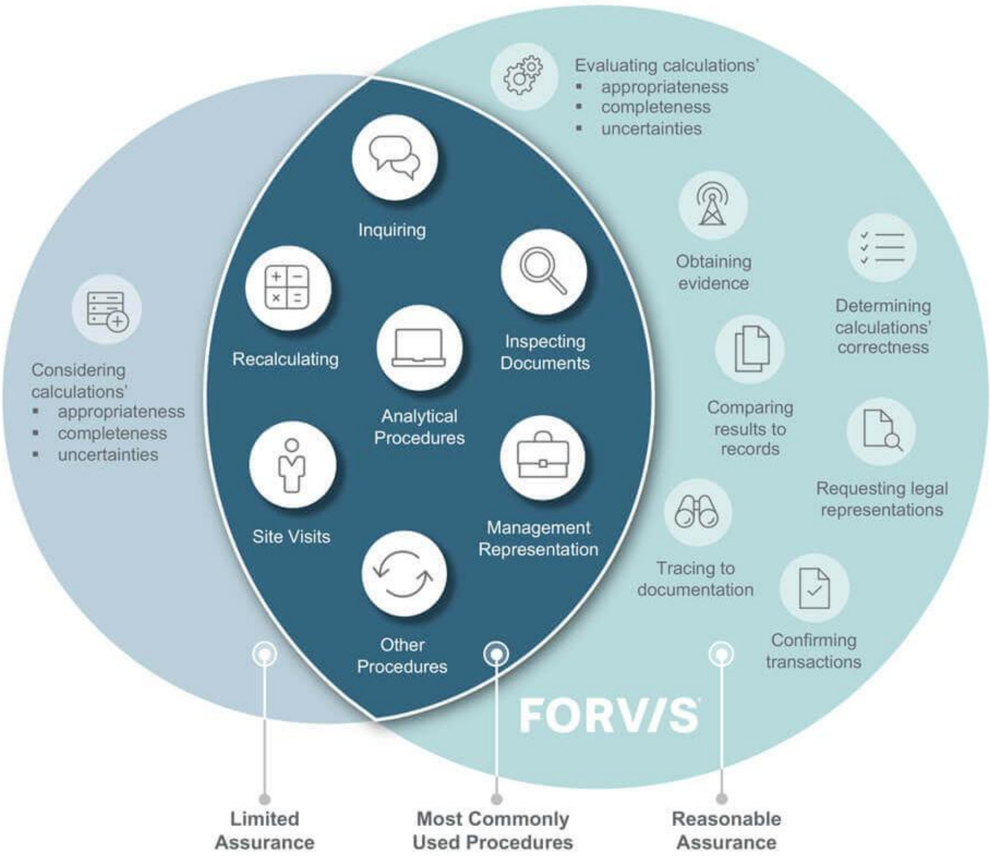

Figure 4: Integrated Framework of AI-Driven ESG Assurance and Greenwashing Detection (Curated by the author)

Conclusion and Implications

Conclusion

This study set out to examine the role of Artificial Intelligence (AI) in strengthening ESG assurance and detecting greenwashing in emerging markets, where sustainability reporting practices are rapidly expanding but institutional assurance mechanisms remain uneven. Drawing on AI-driven textual analysis, anomaly detection, and cross-validation of corporate disclosures with external data sources, the study provides empirical and conceptual evidence that AI can significantly enhance the credibility, reliability, and transparency of ESG reporting.

The findings demonstrate that traditional ESG assurance mechanisms, while important, are insufficient to address the scale, complexity, and narrative-driven nature of contemporary sustainability disclosures, particularly in emerging economies characterized by voluntary reporting regimes and limited enforcement capacity. AI-enabled assurance systems address these limitations by enabling continuous monitoring, multi-source verification, and systematic identification of discrepancies between corporate sustainability claims and underlying performance.

The results confirm that AI-driven ESG assurance is negatively associated with greenwashing practices and positively associated with disclosure credibility and stakeholder trust. Natural language processing techniques are shown to be especially effective in detecting narrative-based greenwashing, which often escapes conventional audit-based verification. Machine learning–based anomaly detection further strengthens assurance by identifying abnormal reporting patterns that may signal symbolic compliance rather than substantive sustainability performance.

Importantly, the study highlights the moderating role of institutional quality, demonstrating that AI-driven assurance mechanisms are most effective when embedded within supportive regulatory and governance environments. Nonetheless, even in weaker institutional contexts, AI contributes meaningfully to greenwashing detection by reducing information asymmetry and expanding oversight capacity.

Overall, the study advances the understanding of ESG assurance by positioning AI not as a substitute for human judgment, but as a complementary governance infrastructure that enhances accountability, transparency, and trust in sustainability reporting systems. By focusing on emerging markets, the research extends ESG and greenwashing scholarship beyond developed-economy settings and responds to growing global demands for technology-enabled sustainability governance.

Theoretical Implications

This study contributes to theory in several important ways. First, it extends greenwashing theory by empirically demonstrating that greenwashing can be operationalized and detected through AI-driven analytical techniques rather than inferred solely through reputational or regulatory outcomes. This shifts greenwashing research from retrospective identification to proactive and scalable detection.

Second, the study advances ESG assurance literature by conceptualizing assurance as a dynamic, data-driven process rather than a static, periodic verification exercise. The integration of AI into assurance frameworks expands the theoretical boundaries of assurance research to include continuous monitoring, narrative analysis, and cross-data triangulation.

Third, from an information asymmetry and signaling perspective, the findings show how AI reduces opportunistic signaling by constraining firms' ability to rely on symbolic sustainability narratives. By increasing the cost of misleading disclosures, AI strengthens the credibility of ESG signals and enhances stakeholder interpretation.

Finally, the study contributes to the emerging discourse on digital governance, highlighting AI's role as an accountability mechanism rather than merely an efficiency-enhancing tool. This reframing is particularly relevant for sustainability governance in emerging markets.

Managerial Implications

For corporate managers, the findings underscore the importance of aligning sustainability communication with verifiable performance outcomes. As AI-driven assurance tools become more prevalent, firms relying on symbolic or selective ESG disclosures face increased detection risk. Managers are therefore encouraged to invest in robust ESG data systems, internal controls, and performance measurement frameworks that support transparent and consistent reporting.

The study also suggests that firms can proactively use AI-based analytics internally to assess disclosure consistency, identify potential greenwashing risks, and improve the quality of sustainability reporting before public release. Such proactive adoption can enhance credibility, reduce reputational risk, and strengthen long-term stakeholder relationships.

Policy and Regulatory Implications

From a policy perspective, the findings highlight the potential of AI-driven ESG assurance to supplement regulatory capacity in emerging markets. Regulators can leverage AI tools to monitor large volumes of sustainability disclosures, prioritize high-risk firms for investigation, and support evidence-based enforcement strategies.

The study also suggests the need for regulatory guidance on AI governance, including transparency, explainability, and ethical use of algorithms in ESG assurance. Developing standardized AI-assisted assurance frameworks can help harmonize disclosure practices and reduce regulatory fragmentation across jurisdictions.

Limitations and Future Research Directions

Despite its contributions, this study is subject to certain limitations. The reliance on secondary data may constrain the precision of greenwashing measurement, and AI-based indicators capture risk signals rather than definitive proof of deceptive intent. Future research could incorporate primary data, experimental designs, or regulatory enforcement outcomes to further validate AI-driven detection models.

Future studies may also explore sector-specific applications, cross-country comparative analyses, and the integration of alternative data sources such as satellite imagery or real-time environmental monitoring. Additionally, ethical and governance challenges related to algorithmic bias, transparency, and accountability warrant deeper examination, particularly in emerging-market contexts.

References

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance, 26*(6), 1315–1344. https://doi.org/10.1093/rof/rfac033

Bissoondoyal-Bheenick, E., Brooks, R., Do, H. X., & Nguyen, T. (2024). ESG rating disagreement and aggregation: Implications for sustainable investing. *Journal of Sustainable Finance & Investment*. Advance online publication. https://doi.org/10.1080/20430795.2023.2291634

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

de Freitas Netto, S. V., Sobral, M. F. F., Ribeiro, A. R. B., & da Luz Soares, G. R. (2020). Concepts and forms of greenwashing: A systematic review. *Environmental Sciences Europe, 32*(19), 1–12. https://doi.org/10.1186/s12302-020-0300-3

Dyer, T., Lang, M., & Stice-Lawrence, L. (2025). Sustainability assurance characteristics and the cost of debt. *Accounting, Organizations and Society*. Advance online publication.

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science, 60*(11), 2835–2857.

https://doi.org/10.1287/mnsc.2014.1984

Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People - An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

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

Hristov, I., Chirico, A., & Appolloni, A. (2022). Sustainability reporting and value creation: The role of AI and digital technologies. *Sustainability, 14*(3), 1–18. https://doi.org/10.3390/su14031423

Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *The Accounting Review, 91*(6), 1697–1724. https://doi.org/10.2308/accr-51383

Lagasio, V. (2024). ESG-washing severity index: A machine learning approach to detect misleading sustainability reporting. *Corporate Social Responsibility and Environmental Management, 31*(1), 1–16. https://doi.org/10.1002/csr.2625

Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. *Organization & Environment, 28*(2), 223–249. https://doi.org/10.1177/1086026615575332

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21. https://doi.org/10.1177/2053951716679679

OECD. (2021). *OECD business and finance outlook 2021: AI in business and finance*. OECD Publishing. https://doi.org/10.1787/edfbca02-en

Shankar, R. (2024). Automated greenwashing detection in Indian corporate sustainability reports using natural language processing. *International Journal of Disclosure and Governance*. Advance online publication. https://doi.org/10.1057/s41310-024-00198-4

Simnett, R., Vanstraelen, A., & Chua, W. F. (2009). Assurance on sustainability reports: An international comparison. *The Accounting Review, 84*(3), 937–967. https://doi.org/10.2308/accr.2009.84.3.937.

❑❑❑