http://dx.doi.org/10.18623/rvd.v23.5225

# AI-ENABLED ESG INTELLIGENCE: A SYSTEMATIC REVIEW OF NLP AND PREDICTIVE ANALYTICS FRAMEWORKS FOR AUTOMATED SUSTAINABILITY REPORTING (2020–2025)

## *INTELIGÊNCIA ESG COM IA: UMA REVISÃO SISTEMÁTICA DAS ESTRUTURAS DE PLN E ANÁLISE PREDITIVA PARA RELATÓRIOS AUTOMATIZADOS DE SUSTENTABILIDADE (2020–2025)*

Article received on: 11/18/2025
Article accepted on: 2/13/2026

**Obaid Ur Rehman Qureshi***
*University of Portsmouth University House Winston, United Kingdom
Orcid: https://orcid.org/0009-0009-0361-7341
laxmi@westernglobaluniversity.us

The authors declare that there is no conflict of interest

**Abstract**
Environmental, social, and governance (ESG) reporting is undergoing a fundamental transition from voluntary narrative disclosure to standardised, mandatory, and audit-ready reporting, driven by the convergence of multiple international and regional regulatory frameworks. This shift has generated substantial demand for AI-enabled ESG intelligence systems capable of processing heterogeneous data sources and producing decision-relevant, traceable, and assurance-ready outputs. However, ESG data remains characterised by significant structural fragmentation, definitional inconsistency across rating methodologies, and the absence of standardised taxonomy mappings conditions that create material barriers to supervised learning approaches reliant on consistent ground-truth labelling [9, 12]. This paper presents a PRISMA 2020-compliant systematic review of AI-based ESG reporting automation research for the period January 2020 to February 2025, encompassing 19 studies selected from 412 records initially retrieved across Scopus, Web of Science, IEEE Xplore, ACM Digital Library, SSRN, and Google Scholar. The review is structured across four functional stages: (i) data acquisition and harmonisation; (ii) NLP-based extraction and classification of ESG statements across environmental, social, and governance disclosure pillars; (iii) predictive analytics for forward-looking risk and disclosure gap forecasting; and (iv) governance, assurance readiness, and reporting design. The regulatory analysis encompasses the EU Taxonomy, CSRD/ESRS, ISSB IFRS S1/S2, and the Saudi Capital Market Authority (CMA) ESG Disclosure Guidelines (2023). Domain-adapted transformer models

***Resumo***
*A divulgação de relatórios ambientais, sociais e de governança (ESG) está passando por uma transição fundamental, passando da divulgação narrativa voluntária para relatórios padronizados, obrigatórios e prontos para auditoria, impulsionada pela convergência de múltiplos marcos regulatórios internacionais e regionais. Essa mudança gerou uma demanda substancial por sistemas de inteligência ESG baseados em IA, capazes de processar fontes de dados heterogêneas e produzir resultados relevantes para a tomada de decisões, rastreáveis e prontos para certificação. No entanto, os dados ESG continuam caracterizados por uma fragmentação estrutural significativa, inconsistência de definições entre metodologias de classificação e a ausência de mapeamentos taxonômicos padronizados — condições que criam barreiras materiais para abordagens de aprendizado supervisionado que dependem de rotulagem consistente de verdades de campo [9, 12]. Este artigo apresenta uma revisão sistemática em conformidade com o PRISMA 2020 sobre pesquisas de automação de relatórios ESG baseadas em IA para o período de janeiro de 2020 a fevereiro de 2025, abrangendo 19 estudos selecionados entre 412 registros inicialmente recuperados no Scopus, Web of Science, IEEE Xplore, ACM Digital Library, SSRN e Google Scholar. A revisão está estruturada em quatro etapas funcionais: (i) aquisição e harmonização de dados; (ii) extração e classificação baseadas em PLN de declarações ESG nos pilares de divulgação ambiental, social e de governança; (iii) análise preditiva para previsão prospectiva de riscos e lacunas de divulgação; e (iv) governança,*

have demonstrated consistent performance improvements in ESG text classification and entity extraction benchmarks, with F1 score gains of 5–15 percentage points over general-purpose language model baselines [14, 16]. Notwithstanding these advances, ensemble and temporal models demonstrate superior performance for ESG risk prediction tasks, particularly when integrating structured ESG metrics with textual sentiment features. Systemic challenges persist, including the absence of verified ground-truth ESG labels, regulatory domain shift, and limited supply chain reporting data. The synthesis produces a deployable seven-step methodological blueprint validated against ESRS, ISSB, CSRD, and Saudi CMA design constraints, with direct applicability to Saudi Vision 2030 sustainability governance, the Public Investment Fund (PIF) portfolio reporting agenda, and Tadawul-listed entity disclosure obligations.

**Keywords:** ESG Reporting. Natural Language Processing. Predictive Analytics. Systematic Review. CSRD. ESRS. ISSB IFRS S1/S2. EU Taxonomy. Supply Chain Logistics. Explainable AI. Assurance Readiness. Vision 2030. Saudi CMA ESG. GCC Sustainability. Arabic NLP. Sovereign Wealth Fund Disclosure.

*prontidão para garantia e projeto de relatórios. A análise regulatória abrange a Taxonomia da UE, CSRD/ESRS, ISSB IFRS S1/S2 e as Diretrizes de Divulgação ESG da Autoridade do Mercado de Capitais da Arábia Saudita (CMA) (2023). Modelos de transformador adaptados ao domínio demonstraram melhorias consistentes de desempenho em benchmarks de classificação de texto ESG e extração de entidades, com ganhos na pontuação F1 de 5 a 15 pontos percentuais em relação às linhas de base de modelos de linguagem de uso geral [14, 16]. Apesar desses avanços, modelos de conjunto e temporais demonstram desempenho superior para tarefas de previsão de risco ESG, particularmente ao integrar métricas ESG estruturadas com características de sentimento textual. Persistem desafios sistêmicos, incluindo a ausência de rótulos ESG de referência verificados, mudanças no domínio regulatório e dados limitados de relatórios da cadeia de suprimentos. A síntese produz um plano metodológico de sete etapas pronto para implementação, validado em relação às restrições de projeto da ESRS, ISSB, CSRD e CMA da Arábia Saudita, com aplicabilidade direta à governança de sustentabilidade da Visão Saudita 2030, à agenda de relatórios da carteira do Fundo de Investimento Público (PIF) e às obrigações de divulgação das entidades listadas na Tadawul.*

***Palavras-chave:*** *Relatórios ESG. Processamento de Linguagem Natural. Análise Preditiva. Revisão Sistemática. CSRD. ESRS. ISSB IFRS S1/S2. Taxonomia da UE. Logística da Cadeia de Suprimentos. IA Explicável. Prontidão para Garantia. Visão 2030. ESG da CMA Saudita. Sustentabilidade do CCG. PLN Árabe. Divulgação de Fundos Soberanos.*

## 1 INTRODUCTION

Over the past five years, ESG reporting has evolved from a reputational disclosure practice into a mandatory, data-intensive governance obligation embedded within multiple converging international regulatory frameworks. For logistics and supply chain-intensive industries, this evolution has extended reporting responsibilities beyond immediate operational boundaries to encompass upstream and downstream supply chain partners across multiple tiers, jurisdictions, and languages, creating acute data collection, harmonisation, and verification challenges. The emergence of AI-enabled reporting infrastructure offers a systematic pathway to address these challenges; however, its design must accommodate not only technical performance requirements but the full set of regulatory, auditability, and governance constraints that mandatory disclosure frameworks impose.

The regulatory trajectory driving this transformation is both accelerating and global. The EU Taxonomy Regulation (EU) 2020/852 established a classification framework for environmentally sustainable economic activities [3]. The Corporate Sustainability Reporting Directive (EU) 2022/2464 (CSRD) substantially expanded mandatory ESG reporting obligations and introduced limited assurance requirements [2]. The European Sustainability Reporting Standards (ESRS), introduced as a delegated regulation in 2023, codify the disclosure schema across general and topical standards [4]. At the global level, IFRS S1 and IFRS S2, issued by the International Sustainability Standards Board (ISSB) in June 2023, establish the baseline sustainability disclosure framework governing governance, strategy, risk management, and metrics and targets [5]. The EU Corporate Sustainability Due Diligence Directive (CSDDD) (EU) 2024/1760 mandates value chain due diligence on environmental and human rights impacts, with direct implications for logistics operators [8]. In the United States, the SEC adopted climate disclosure rules (Release 33-11275) in March 2024; however, these rules are subject to ongoing judicial review following a partial implementation stay granted by the U.S. Court of Appeals (Eighth Circuit) in April 2024, and their final operational status remains unresolved as of early 2026 [6, 7].

Beyond the EU and US, GCC regulatory authorities have established parallel ESG disclosure frameworks of direct relevance to AI-enabled reporting automation. The Saudi Capital Market Authority (CMA) issued mandatory ESG Disclosure Guidelines in 2023, applicable to all Tadawul-listed entities, requiring structured disclosure in both Arabic and English [A]. The Saudi Exchange launched the Saudi ESG Index (SASESG) to benchmark listed entity ESG performance [B]. The UAE Securities and Commodities Authority issued a sustainability reporting framework in 2023 [E]. Saudi Arabia’s Vision 2030 Green Initiative targets 50% renewable energy by 2030 and net-zero emissions by 2060 [D], while the PIF ESG Investment Policy governs sustainability disclosure across the sovereign wealth fund's portfolio companies [C]. These GCC frameworks impose AI design constraints analogous to ESRS and ISSB but with additional requirements for Arabic-language output, Shariah governance considerations, and proportionate reporting provisions for emerging market entities.

This convergence of regulatory obligations has transformed ESG reporting into a quasi-regulated information system [14]. The disclosure fragmentation challenge remains acute: ESG rating agencies diverge substantially in measurement approaches and metric weightings, generating what Berg *et al*. [9] term "aggregate confusion" a structural inconsistency that creates fundamental barriers to supervised learning. The data heterogeneity constraint is further compounded by different disclosure approaches across jurisdictions, industries, and entity sizes. For ESG reporting practitioners, this dual challenge motivates growing interest in AI-based solutions capable of automating, standardising, and assurance-enabling ESG disclosure at enterprise scale.

The evidence base supporting AI-enabled ESG reporting has strengthened considerably over the review period. Recent studies demonstrate that NLP models can effectively extract ESG information across subdomains, outperforming traditional bag-of-words approaches in classification precision and disclosure coverage [14]. Finance-centric transformer models such as FinBERT have established competitive performance on financial text classification tasks, providing a validated baseline for ESG disclosure and sentiment analysis [13]. Text-based automated grading systems demonstrate the feasibility of mapping large document corpora to ESG categories in real time [15]. Predictive analytics extends this capability to forward-looking risk identification, disclosure gap forecasting, and controversy prediction from integrated data streams.

While related reviews by Meng *et al*. [26] and Elhady and Shohieb [27] address NLP approaches for sustainability reporting and computational ESG tools respectively, the present review makes three distinct contributions. First, it provides the first PRISMA-structured synthesis that integrates both NLP pipeline design and predictive analytics frameworks for logistics-sector ESG automation within a single review. Second, it proposes a deployable seven-step methodological blueprint validated against ESRS, ISSB, CSRD, and Saudi CMA design constraints, with explicit Arabic-language NLP requirements and Vision 2030-aligned deployment contexts. Third, it establishes a governance and assurance readiness framework with specific traceability and auditability specifications applicable to both EU-regulated and GCC-regulated entities.

## 2 AIM AND OBJECTIVES OF THE STUDY

Aim of the Study: This study aims to synthesise the state of evidence on AI-enabled ESG reporting automation (January 2020–February 2025), identify dominant technical approaches and their performance boundaries, characterise governance and assurance constraints on enterprise deployment, and produce a deployable methodological blueprint for logistics and supply chain organisations operating under ESRS, ISSB, CSRD, and Saudi CMA reporting obligations.

Objectives of the Study:

(1) Define and map ESG report automation use cases across the full reporting lifecycle from data acquisition and NLP-based extraction to predictive analytics and final disclosure assurance with reference to logistics and supply chain operational contexts.

(2) Define, compare, and evaluate the major NLP approaches applied to ESG reporting automation specifically dictionary-based, classifier-based, transformer-based, and retrieval-augmented generation (RAG) methods with respect to performance characteristics, limitations, and suitability across ESG criteria, industry contexts, and GCC bilingual reporting requirements.

(3) Define, compare, and evaluate the major predictive analytics approaches applied to ESG risk and disclosure forecasting specifically ensemble, temporal, and controversy prediction model families with respect to forecast targets, validation requirements, and operational implications for Vision 2030-aligned logistics management.

(4) Characterise AI design constraints imposed by emerging global and regional regulatory requirements, including EU Taxonomy, CSRD/ESRS, ISSB IFRS S1/S2, SEC Climate Rules, and Saudi CMA ESG Guidelines, with specific attention to taxonomy alignment, double materiality calibration, traceability, auditability, and proportionality.

(5) Develop a deployable, audit-ready seven-step ESG reporting automation blueprint applicable to logistics and supply chain organisations, validated against ESRS, ISSB, CSRD, and Saudi CMA ESG design constraints, and aligned with Vision 2030 sustainability governance objectives.

## 3 REVIEW METHODOLOGY

### 3.1 Design and reporting standards

This study employs a structured systematic review design, with reporting following PRISMA 2020 standards for transparent and reproducible systematic reviews [1]. Database searches were conducted on 28 February 2025, covering literature published between January 2020 and February 2025 across Scopus, Web of Science, IEEE Xplore, ACM Digital Library, SSRN, and Google Scholar. Search terms combined ESG reporting and disclosure concepts with AI technique identifiers including: ‘NLP’; ‘transformer’; ‘BERT’; ‘FinBERT’; ‘knowledge graph’; ‘topic modeling’; ‘predictive analytics’; ‘XGBoost’; ‘time series’; ‘explainable AI’; and ‘retrieval augmented generation’. Regulatory and standards documents were reviewed in parallel [2–8, A–F].

### 3.2 PRISMA 2020 article selection

Table 1 presents the full PRISMA 2020-compliant article selection summary. The final synthesis corpus of 19 studies was selected from 412 database records initially retrieved. Studies were selected based on their direct contribution to AI-enabled ESG reporting automation architectures, NLP pipeline design, predictive analytics applications, or governance and assurance frameworks for ESG AI systems.

**Table 1**

*PRISMA 2020 article selection summary.*

| PRISMA 2020 Stage | Process Description | N (Records) |
|---|---|---|
| Records identified via database searches | Scopus, WoS, IEEE Xplore, ACM Digital Library, SSRN, Google Scholar | 412 |
| Records identified via supplementary methods | Citation tracking; regulatory document repositories | 24 |
| Records after duplicate removal | Automated deduplication followed by manual verification | 299 |
| Records screened (title and abstract review) | Inclusion/exclusion criteria applied | 299 |
| Records excluded at screening stage | Off-topic; non-AI/NLP method; outside date range; non-English | 231 |
| Full-text articles assessed for eligibility | Full-text content assessment against all criteria | 68 |
| Full-text articles excluded (with reason) | Insufficient technical detail (31); no ESG context (18) | 49 |
| Studies included in qualitative synthesis | Final corpus informing all substantive review sections | 19 |

Database searches conducted 28 February 2025.

### 3.3 Eligibility criteria

Inclusion criteria required each article to: (a) be published between January 2020 and February 2025; (b) propose, apply, or review AI or NLP techniques for ESG reporting, disclosure, or sustainability scoring; and (c) provide sufficient methodological information to assess data inputs, model architecture, and evaluation approach. Exclusion criteria eliminated articles lacking: (a) sufficient technical detail regarding model design or evaluation protocol; (b) contextual grounding in ESG or sustainability reporting; or (c) an empirical or methodological substantiation (i.e., conference abstracts, editorials, or opinion pieces were excluded).

### 3.4 Data extraction and synthesis protocol

Data were extracted across seven standardised dimensions: (a) dataset and setting; (b) ESG taxonomy and labelling scheme applied; (c) NLP technique employed; (d) predictive analytics model applied, where relevant; (e) evaluation metrics reported; (f) explainability and traceability mechanisms present; and (g) deployment context and operational implications. Given significant methodological heterogeneity across included studies encompassing different datasets, task definitions, evaluation metrics, and NLP architectures direct quantitative meta-analysis was not feasible. A structured qualitative synthesis was therefore employed, consistent with established practice in related AI and sustainable finance review literature [26, 27].

## 4 THE REGULATORY AND STANDARDS ENVIRONMENT AS A DESIGN CONSTRAINT FOR ESG AUTOMATION

AI-enabled ESG intelligence systems must be designed for regulatory reportability the capacity to produce outputs that satisfy specific disclosure regime requirements, withstand assurance scrutiny, and remain valid as standards evolve. Table 2 summarises the principal regulatory frameworks and their primary AI design constraint implications.

**Table 2**

*Principal ESG regulatory frameworks and AI design constraints. †SEC rules subject to judicial review and partial implementation stay as of early 2026 [6, 7].*

| Framework | Jurisdiction | Primary Scope | Key AI Design Constraints | Ref. |
|---|---|---|---|---|
| EU Taxonomy (2020/852) | EU | Environmental activity classification | Taxonomy alignment; machine-readable activity mapping | [3] |
| CSRD (2022/2464) | EU | Mandatory ESG reporting + limited assurance | Traceability; auditability; double materiality | [2] |
| ESRS Delegated Reg. (2023) | EU | Disclosure schema and topical standards | Completeness; consistency; proportionality | [4] |
| ISSB IFRS S1/S2 (2023) | Global | Sustainability disclosure baseline | All seven constraints; double materiality | [5] |
| SEC Climate Rules (2024)† | USA | Climate disclosure for listed issuers | Taxonomy alignment; consistency | [6,7] |
| CSDDD (2024/1760) | EU | Supply chain due diligence | Traceability; auditability; evidence management | [8] |

| Framework | Jurisdiction | Primary Scope | Key AI Design Constraints | Ref. |
|---|---|---|---|---|
| Saudi CMA ESG Guidelines (2023) | Saudi Arabia / GCC | ESG disclosure for Tadawul-listed entities | All seven + Arabic/English bilingual output | [A] |
| UAE SCA Framework (2023) | UAE / GCC | Sustainability reporting: UAE-listed entities | Completeness; consistency | [E] |
| GCC Exchanges Guidance (2022) | GCC | Listed entity ESG metrics baseline | Taxonomy alignment; consistency | [F] |

### 4.1 International and eu regulatory frameworks

The EU Taxonomy Regulation (EU) 2020/852 establishes the classification framework for environmentally sustainable activities, imposing taxonomy alignment as the primary AI design requirement [3]. The CSRD (EU) 2022/2464 substantially expands mandatory reporting obligations and introduces limited assurance requirements, demanding AI systems capable of producing traceable, evidence-grounded, auditable outputs [2]. The ESRS delegated regulation (2023) defines the detailed disclosure schema against which completeness and consistency of AI-generated disclosures are evaluated [4]. IFRS S1 and IFRS S2, issued by the ISSB in June 2023, establish the global baseline for sustainability disclosure [5].

These frameworks collectively impose seven AI system design constraints: taxonomy alignment, traceability, completeness, consistency, auditability, materiality calibration, and proportionality. Two constraints merit particular emphasis for their architectural implications. First, materiality calibration: both ESRS and ISSB require double materiality assessment, distinguishing impact materiality (effects of the entity on people and environment) from financial materiality (effects on enterprise value), necessitating dual-output classification architectures. Second, proportionality: CSRD and ISSB allow simplified reporting for SMEs and value chain participants, requiring AI output granularity to be configurable by entity type and reporting obligation scope.

### 4.2 GCC and vision 2030 regulatory landscape

The GCC has established a parallel and rapidly maturing ESG disclosure ecosystem. The Saudi CMA issued mandatory ESG Disclosure Guidelines in 2023, requiring structured disclosure in both Arabic and English for all Tadawul-listed entities [A]. The Saudi Exchange launched the Saudi ESG Index (SASESG) [B]. The UAE SCA issued a sustainability reporting framework in 2023 [E], and GCC Exchanges published collective ESG reporting guidance [F].

Saudi Arabia's Vision 2030 sustainability architecture further shapes the AI design environment. The Saudi Green Initiative targets 50% renewable energy by 2030 and net-zero by 2060 [D], creating specific Scope 1, 2, and 3 data management requirements. The PIF ESG Investment Policy governs disclosure across the sovereign wealth fund's portfolio companies [C]. The National Logistics Strategy (2021) and Saudi Ports Authority sustainability mandates generate direct operational demand for logistics-sector ESG automation aligned with the architectures developed in this review. These GCC frameworks introduce additional constraints for Arabic-language processing, Shariah governance considerations, and proportionate reporting for emerging market entities.

## 5 DATA FOUNDATIONS FOR ESG INTELLIGENCE

### 5.1 Data types and collection

ESG reporting automation requires integration of both structured data (emissions inventories, energy consumption, occupational safety statistics, diversity metrics, supplier performance data, and financial risk indicators) and unstructured data (sustainability reports, management discussions, corporate website content, policy documents, supplier declarations, news articles, and social media). The analytical boundary between these data types is increasingly porous as AI extraction systems enable structured information retrieval from unstructured source documents.

ESG data remains characterised by significant structural fragmentation, definitional inconsistency across rating methodologies, and the absence of standardised taxonomy mappings conditions that create material barriers to supervised learning approaches [9, 12]. Major inconsistencies in metric definitions across rating agencies generate substantial economic and analytical consequences [9, 11, 12]. The absence of reliable ground truth requires ESG AI systems to employ alternative supervision strategies, including weak supervision frameworks, multi-rater aggregation, and regulatory-anchored labelling schemes, each carrying validity limitations that must be explicitly acknowledged in system design.

## 5.2 Supply chain and logistics data challenges

Logistics-intensive organisations typically carry disproportionately high Scope 3 emissions exposure often exceeding 70% of total GHG footprint because their value chain emissions are distributed across geographically and operationally fragmented supplier networks spanning multiple tiers, transport modes, jurisdictions, and languages [31, 32]. The CSDDD (EU) 2024/1760 imposes due diligence obligations requiring supplier-level ESG data at granularity that traditional periodic reporting systems cannot support [8]. The linguistic diversity, format heterogeneity, and verification complexity of international supplier documentation make document ingestion and information extraction central architectural requirements for logistics-oriented ESG intelligence systems.

## 5.3 Harmonisation and knowledge layers

Ahmed *et al*. [19] demonstrate that an intermediate knowledge layer storing extracted ESG entities, metrics, and claims with source document links, extraction timestamps, taxonomy tags, and version identifiers substantially improves system robustness and enables recalculation when standards change under new ESRS interpretations or CMA guideline updates. This architecture also provides the provenance foundation for ESG assurance, enabling narrative claims to be traced to specific source documents and extraction events. The ESG-CID framework [19], employing disclosure content indices with weak supervision, further demonstrates the feasibility of building retrieval-augmented systems capable of processing long-form ESG reports without losing traceability. This is particularly relevant for Vision 2030-aligned deployments where

Arabic and English versions must be processed in parallel with consistent ESG classification outputs.

## 6 NLP FOR ESG REPORTING AUTOMATION

Figure 1 presents the proposed multi-layer ESG intelligence pipeline for automated, audit-ready sustainability reporting, integrating five functional layers: (1) data ingestion; (2) NLP-based classification and entity extraction; (3) knowledge graph harmonisation and taxonomy alignment; (4) predictive analytics; and (5) assurance-ready reporting output with embedded traceability and explainability controls.

Figure 1. Proposed multi-layer ESG intelligence pipeline for automated, audit-ready sustainability reporting. Layer 1: Data ingestion; Layer 2: NLP classification and entity extraction; Layer 3: Knowledge graph harmonisation; Layer 4: Predictive analytics; Layer 5: Assurance-ready output with provenance traceability. [Insert final authored figure at production stage.]

### 6.1 Baselines: dictionaries, topic models, and classical classifiers

Early approaches to ESG automation relied on dictionary-based methods and topic models for extracting ESG-relevant content from disclosure documents. While providing analytical transparency with no labelled training data requirements, they underperform on contextually nuanced and domain-varied ESG language, particularly across multiple ESG subdomains or regulatory frameworks [14, 16]. Supervised classical machine learning models including Support Vector Machines and Logistic Regression improve classification precision with increasing labelled data volumes but exhibit material performance degradation under domain shift, for instance when transitioning from annual sustainability reports to news articles or across industry sectors with substantially different ESG vocabularies.

### 6.2 Transformer models and domain adaptation

The majority of research from 2020 to 2025 employs transformer-based architectures, motivated by their capacity to model long-range contextual dependencies and discourse-level nuances in ESG narrative text. FinBERT demonstrated strong benchmark performance on financial text classification and sentiment analysis tasks, establishing a relevant and widely-cited baseline for ESG disclosure analysis [13]. However, subsequent ESG-specific pre-training approaches have demonstrated performance advantages on ESG subdomain classification tasks [14], and more recent instruction-tuned large language models (including domain-adapted GPT-4 and Llama-3 variants) have demonstrated competitive results on specific ESG extraction and classification subtasks, indicating the field has advanced beyond FinBERT as a universal ESG benchmark. Domain adaptation techniques continued pre-training on ESG corpora and task-specific fine-tuning consistently improve classification performance in ESG shared task benchmarks [16, 17].

### 6.3 ESG-specific pretraining and subdomain models

A significant methodological advance in the review period is ESG-specific pre-training, in which separate language models are pre-trained on large ESG report and news article corpora for each disclosure pillar (E, S, G) prior to fine-tuning on downstream classification tasks. Schimanski *et al*. [14] demonstrate that this approach achieves improved subdomain classification across 13.8 million texts, maintaining model explainability features practically relevant for assurance contexts. The pillar-specific strategy reflects the genuinely distinct vocabulary, regulatory basis, and measurement approaches associated with environmental, social, and governance disclosures.

### 6.4 From classification to extraction: entities, metrics, and claims

ESG reporting automation requires more than document-level topic classification. To generate disclosure-grade outputs satisfying ESRS, ISSB, or CMA requirements, AI systems must extract specific entities (facilities, suppliers, geographic locations, certifying bodies), quantitative metrics (absolute and intensity-based emissions values, injury rates, energy consumption, diversity percentages), and structured claims (policy commitments, reduction targets and their time horizons). Recent studies have applied sequence labelling, relation extraction, and template filling to this challenge. For logistics-sector applications, entities of particular importance include transport emissions classifications, fleet composition data, transport mode specifications, carrier-level Scope 3 data, and supplier compliance declarations.

### 6.5 Retrieval-augmented generation and long documents

ESG annual reports typically span 100 to 400 pages [33], creating fundamental challenges for transformer models under context window constraints. Retrieval-Augmented Generation (RAG) architectures address this by grounding model outputs in retrieved document segments rather than parametric model memory, maintaining the source-document traceability required by assurance frameworks. Ahmed *et al*. [19] propose ESG-CID a dataset and methodology employing disclosure content indices with weak supervision to assist ESG-oriented language models in satisfying RAG constraints specific to long-form ESG document automation demonstrating improved retrieval precision and traceable output generation. This approach is particularly relevant for Vision 2030-aligned deployments where Arabic and English report versions must be processed consistently.

### 6.6 Evaluation beyond accuracy: faithfulness and auditability

While standard classification metrics such as F1 score and AUC-ROC remain necessary baseline measures for benchmarking ESG NLP model performance, they are insufficient as sole evaluation criteria for regulatory-grade ESG reporting automation. These metrics do not capture disclosure faithfulness (whether AI output accurately reflects source evidence), coverage completeness (whether all mandatory disclosure items are addressed), or auditability (whether outputs can be traced to verified source documents). Emerging research reframes ESG output evaluation as an information quality task: does the generated output satisfy all disclosure elements as specified by the applicable standard? This framing has direct implications for how AI systems should be benchmarked, certified, and integrated into formal assurance workflows.

### 6.7 Practical prompting and control mechanisms for report drafting

Where generative models are employed for narrative ESG report drafting, the primary engineering challenge lies in constraining outputs to evidence-grounded content, preventing hallucinated or unsupported claims, and ensuring disclosure consistency with the underlying structured data layer. The “retrieve, then draft, then verify” architecture addresses this systematically: (i) retrieval models identify relevant evidence passages from the knowledge layer; (ii) the generative model drafts only using retrieved content; and (iii) post-generation verification checks numerical consistency, terminology alignment, boundary assumptions, and internal contradictions.

For logistics-sector emissions reporting, narrative control is critical regarding emissions calculation methodology disclosures, including market-based versus location-based electricity accounting, transport mode-specific emission factor selection, and cargo allocation methodology choices. In a well-governed knowledge layer architecture, all such methodological assumptions are stored as metadata, linked to specific narrative passages, and available for auditor verification directly supporting CSRD assurance requirements, ISSB disclosure standards, and Saudi CMA traceability obligations.

### 6.8 Multilingual and cross-jurisdiction NLP considerations

Cross-border supply chains generate ESG documentation across multiple languages and regulatory frameworks. NLP systems must accommodate cross-lingual concept alignment while managing the risk that machine translation of domain-specific regulatory terminology introduces systematic errors. A hybrid architecture employing multilingual models for routing and domain-specific models for extraction, with text aligned to a shared regulatory concept layer offers the most architecturally sound and technically reliable approach. Systems must also accommodate structural differences between disclosure frameworks: EU-centric regimes (CSRD/ESRS, EU Taxonomy) mandate schema-structured outputs with specific data point requirements, while other jurisdictions allow greater narrative flexibility.

A critical and underserved cross-jurisdictional challenge is Arabic-language ESG reporting. The Saudi CMA mandates bilingual Arabic and English ESG disclosures for all Tadawul-listed entities [A], creating direct operational requirements for Arabic-language ESG NLP. Current Arabic NLP infrastructure lacks domain-adapted ESG models comparable to FinBERT, and publicly available Arabic ESG-labelled corpora are extremely limited. Arabic presents processing challenges absent from English ESG NLP, including high morphological complexity, right-to-left directionality, and dialectal variation across GCC jurisdictions. This infrastructure gap represents both a significant research priority and a practical deployment barrier for Vision 2030-aligned ESG automation, particularly for Saudi logistics operators, Aramco supply chain reporting, and PIF portfolio company disclosures.

## 7 PREDICTIVE ANALYTICS FOR ESG RISK AND DISCLOSURE FORECASTING

### 7.1 The strategic role of ESG prediction

Predictive analytics extends the function of ESG reporting beyond retrospective compliance documentation, transforming it into a forward-looking governance and risk management instrument. This is of particular strategic value in logistics network management, where early identification of supplier risk exposure, carbon intensity trajectories, and disclosure gap forecasts enables proactive ESG governance. For Vision 2030-aligned logistics operators including Saudi Ports Authority entities, NIDLP participants, and NEOM supply chain partners the capacity to forecast ESG risk events and compliance gaps directly supports operational planning and national sustainability target-setting commitments.

## 7.2 Model families and feature design

Studies published between 2020 and 2025 consistently favour multimodal feature designs combining structured ESG metrics with textual features derived from news sentiment and controversy monitoring. Ensemble models particularly gradient boosting and XGBoost variants demonstrate consistent and reliable performance on ESG risk prediction tasks due to their capacity to handle missing values and model non-linear feature interactions [15]. Temporal models, including LSTM-based architectures and temporal convolutional networks, are particularly effective where features are structured around time-series news archives, as demonstrated by Lee *et al*. [15]. Sentiment-based models employing FinBERT with nested sentiment architectures [18] highlight the importance of model interpretability tools specifically SHAP attribution when sentiment features serve as predictive signals for ESG market dynamics, a finding directly relevant for regulatory deployment where explainability is a governance requirement.

## 7.3 Forecast targets and use cases

Four distinct forecast targets have been identified in the reviewed ESG prediction literature:

- ESG score and risk prediction: forecasting entity-level ESG ratings or composite risk scores for the next reporting period, enabling anticipatory remediation strategies and proactive investor communication;
- Financial materiality forecasting: estimating the financial performance implications of specific ESG risk factors, including impacts on cost of equity and capital market access directly relevant to the ISSB double materiality assessment requirement;
- Disclosure gap forecasting: predicting which mandatory disclosure elements are likely to be incomplete or absent based on current data availability a function of high practical value for pre-submission assurance workflows and regulatory readiness assessments; and

- Controversy and reputational event prediction: early identification of emerging ESG-related media controversies, stakeholder disputes, or regulatory enforcement events from news and social media signals, enabling proactive risk management prior to materialisation as ESG rating downgrades or investor relations events [18]. For logistics operators, supplier controversy prediction serves as a leading risk indicator for due diligence obligations under CSDDD [8].

### 7.4 Validation and robustness

ESG prediction model validation faces three structural challenges: regime shifts (emergence of new regulatory standards), label drift (rating agency methodology changes), and domain shift (industry-specific vocabulary evolution). The ESG rating disagreement literature establishes that single-rater labelled datasets do not generalise across investors and rating methodologies [9, 11, 12], requiring robustness testing against multiple alternative label sources. For reliable regulatory deployment, ESG prediction systems should incorporate temporal holdout validation, cross-sector testing, sensitivity analysis for alternative label sources, and calibration reporting for risk score thresholding applications.

### 7.5 Integrating predictive signals with operational planning

The operational value proposition of ESG predictive analytics for logistics managers lies in translating ESG signals into actionable operational intelligence: predicting carbon intensity trajectories across route and modal alternatives, forecasting warehouse energy requirements under seasonal demand variation, or identifying supplier non-compliance trajectories before they generate service-level or regulatory consequences. When integrated into logistics planning systems, ESG analytics shifts from periodic compliance reporting to continuous sustainability performance management a transformation aligned with Vision 2030's emphasis on data-driven national logistics infrastructure development.

The concept of scenario-aware reporting offers a further strategic innovation: AI systems generating prospective emissions scenarios (e.g., under current fleet composition versus progressive electrification investment paths) alongside historical reporting, enabling forward-looking sustainability commitments within formal disclosure structures. Appropriate disclaimers must clearly distinguish modelled trajectories from binding commitments, with versioned data provenance links embedded in the report to satisfy assurance requirements. This capability supports evidence-based sustainability target-setting and transparent stakeholder communication, extending ESG reporting well beyond annual regulatory compliance.

**Table 3**

*Representative 2020–2025 studies informing AI-enabled ESG reporting automation.*

| Study | Dataset / Setting | Primary Task | Method | Key Contribution | XAI / Traceability | Implication for Reporting Automation |
|---|---|---|---|---|---|---|
| Berg *et al*. (2022) [9] | Multiple rating agencies; cross-firm | Rating divergence analysis | Decomposition analysis | Quantifies divergence by scope, measurement, and weighting methodology | N/A | Establishes ground-truth instability as fundamental barrier to supervised ESG learning |
| Schimanski *et al*. (2024) [14] | 13.8M texts (reports and news) | ESG subdomain classification | Pre-trained ESG NLP models | Improved subdomain classification at scale with model explainability features | Model explainability discussed | Strong backbone architecture for ESG extraction and scoring pipelines |
| Huang *et al*. (2023) [13] FinBERT | Financial corpora | Financial-text NLP baseline | Transformer (FinBERT) | Superior performance on financial sentiment and NLI over general-purpose BERT | Attention-based | Validated baseline for ESG disclosure and sentiment analysis |
| Lee *et al*. (2024) [15] | News archive (LexisNexis) | Automated ESG grade assessment | Text classification with weak supervision | Automates E/S/G grading pipeline from real-time news feeds | Limited | Enables continuous operational ESG scoring from news streams |
| Goel *et al*. (2022) [16] | FinNLP ESG shared task | ESG sentence classification | BERT with linguistic features | Competitive performance with hybrid feature- | Limited | Highlights feature engineering value alongside transformer fine-tuning |

| Study | Dataset / Setting | Primary Task | Method | Key Contribution | XAI / Traceability | Implication for Reporting Automation |
|---|---|---|---|---|---|---|
| | | | | transformer architecture | | |
| Mashkin *et al*. (2023) [17] | FinNLP ML-ESG task | Transformer benchmarking | Transformer model variants | Systematic benchmarks for ESG classification across multiple model families | Limited | Establishes comparative evaluation practices for ESG NLP model selection |
| Saxena *et al*. (2024) [18] | News + product perception | Nested ESG sentiment; controversy prediction | FinBERT + SHAP explainability | High-accuracy sentiment classification with controversy prediction and SHAP attribution | XAI included (SHAP) | Demonstrates explainability of sentiment signals for reputational risk forecasting |
| Ahmed *et al*. (2025) [19] | ESG annual reports (long documents) | Retrieval for ESG language models | RAG with disclosure content index (ESG-CID) | Improves retrieval precision via weak supervision for long-form ESG documents | Grounded retrieval with traceability | Supports traceable, audit-ready ESG narrative generation with provenance |

Source: Author synthesis from included studies. XAI = Explainable Artificial Intelligence; RAG = Retrieval-Augmented Generation; SHAP = SHapley Additive exPlanations.

## 8 GOVERNANCE, EXPLAINABILITY, AND ASSURANCE READINESS

### 8.1 Traceability as a first-class system requirement

Mandatory ESG assurance both limited and reasonable assurance as progressively required under CSRD and ISAE 3000 requires that every material claim and quantitative figure can be traced to a specific source document, an identified extraction process, and a responsible data custodian. In the context of AI-generated ESG outputs, this demands a complete provenance chain: the capacity to identify, for any disclosed figure or narrative statement, the source document version, the NLP extraction pathway, the knowledge layer storage event, and the human review decision that authorised its inclusion.

Knowledge graph layers and RAG pipelines provide the architectural foundation for such provenance [19]. In logistics contexts, this extends to supplier-level claim documentation, where third-party declarations of environmental performance must be linkable to specific disclosure statements and their data sources.

## 8.2 Explainability and multi-stakeholder trust

Explainability in ESG AI systems is not a single technical specification but a multi-stakeholder governance obligation. Internal audit teams require evidence extraction paths and model confidence scores. Sustainability teams require taxonomy classification rationale and disclosure completeness explanations. Legal teams require disclosure boundary justifications. Assurance providers require full provenance records and methodology documentation. Feature importance tools (SHAP, LIME), evidence snippet extraction, and taxonomy mapping visualisations each serve distinct stakeholder needs within the same AI system. In contentious disclosure scenarios particularly Scope 3 boundary decisions or CSDDD due diligence classification definitional transparency and methodology documentation function as material governance controls.

## 8.3 Bias, greenwashing, and model misuse

As demonstrated in the ESG rating divergence literature [9, 12], AI models trained predominantly on large-capitalisation, developed-market companies systematically disadvantage smaller firms and entities from data-sparse regions, generating structural classification bias. In supply chain contexts, smaller and emerging-market suppliers who lack comprehensive ESG documentation are disproportionately misclassified as high-risk, creating systematic false positives that undermine due diligence-driven ESG intelligence and generate unjustified supply chain disruptions [8]. The risk of AI-generated ESG outputs being deployed as superficial compliance box-ticking producing narrative appearances of disclosure without verifiable evidential grounding must be addressed through output verification requirements and human-in-the-loop approval gates.

### 8.4 Continuous monitoring and standard drift

ESG AI systems require ongoing monitoring across three dimensions: (i) input data distribution shifts, which may reflect changes in corporate reporting practices or data source availability; (ii) taxonomy version changes, which may invalidate previously generated classification labels and require systematic relabelling; and (iii) output completeness changes, reflecting additions of new mandatory disclosure requirements in updated regulatory standards. A versioning and revalidation approach, aligned with annual regulatory review cycles including ESRS updates and periodic CMA guideline revisions, provides the most operationally sustainable governance framework for managing cumulative standard drift.

### 8.5 Assurance-oriented metrics for ESG AI systems

As ESG assurance requirements mature under CSRD and analogous GCC frameworks, AI systems require evaluation against metrics designed specifically for assurance contexts. Table 4 defines three proposed assurance-oriented quality metrics.

**Table 4**

*Proposed assurance-oriented quality metrics for ESG AI system evaluation.*

| Metric | Definition | Measurement Approach |
|---|---|---|
| Evidence Coverage | Percentage of mandatory disclosure requirements for which at least one traceable evidence source with a verifiable extraction path has been identified | Automated completeness check against the disclosure requirement graph (Step 1 of the blueprint); periodic human sampling verification |
| Evidence Precision | Percentage of AI-identified evidence items that can be validated through human review or cross-referencing against a structured system of records | Sampling-based audit procedures analogous to financial audit evidence testing; comparison against verified primary data sources |
| Narrative Faithfulness | Percentage of AI-generated narrative content that can be matched to specific evidence spans without hallucinated, unsupported, or inaccurately attributed claims | Automated evidence-span matching with provenance links; human review of low-confidence passages; contradiction detection pipeline output |

These metrics can be operationalised through sampling-based audit procedures analogous to financial audit evidence testing. AI systems capable of reporting these metrics would substantially reduce the friction and cost of third-party ESG assurance engagements by enabling assurance providers to focus manual verification on areas flagged for low evidence coverage or narrative faithfulness.

Implementation guidance: The most resilient automation strategies begin with constrained scope a single reporting standard, a defined business unit, and a bounded set of high-confidence metrics and scale progressively as data quality, governance infrastructure, and stakeholder trust mature. This phased approach is recommended particularly for first-deployment contexts such as Tadawul-listed logistics operators implementing Saudi CMA ESG disclosure requirements.

## 9 SYNTHESIS FOR LOGISTICS AND SUPPLY CHAIN CONTEXTS

ESG intelligence presents particularly high strategic value for logistics and supply chain industries, given their combination of high Scope 3 emissions exposure, extensive supplier network dependencies, mandatory due diligence obligations under CSDDD [8], and growing regulatory exposure through EU and GCC supply chain disclosure requirements. The NLP capabilities reviewed in Section 6 can be applied to systematic ingestion of supplier declarations, audit reports, contracts, and policy documents, enabling automated extraction of supplier ESG commitments, performance statements, and identified risk indicators. The predictive analytics capabilities reviewed in Section 7 enable assessment of supplier engagement trajectories and early-warning controversy signals before escalation to CSDDD-reportable or CMA-notifiable events.

Three integrated deployment pipeline architectures are identified as most operationally relevant for logistics organisations:

- Document-to-Template Pipeline: Ingests supplier policies, audit reports, and internal ESG data; applies NLP extraction and classification; and generates ESRS/ISSB/Saudi CMA-compliant disclosure templates with embedded provenance links. This pipeline directly addresses structured reporting obligations of Tadawul-listed logistics entities under 2023 CMA guidelines.
- Signal Triangulation Pipeline: Integrates internal operational data with external signals including news sentiment, regulatory filings, and controversy indicators; applies ensemble and temporal predictive models to identify emerging ESG risks;

and generates risk intelligence for procurement, compliance, and due diligence teams.

- Disclosure Gap Management Pipeline: Maintains a machine-readable, version-controlled graph of disclosure requirements across applicable regulatory regimes; applies gap prediction models to identify missing or insufficient disclosures; and generates prioritised remediation plans calibrated by materiality level.

In the context of Saudi Arabia's Vision 2030 logistics transformation encompassing the National Logistics Strategy (2021), Saudi Ports Authority sustainability mandates, and NEOM's zero-carbon supply chain commitments these pipeline architectures have direct deployment applicability. The document-to-template pipeline supports Aramco supply chain ESG disclosure under CMA guidelines; the signal triangulation pipeline enables controversy detection for PIF portfolio company monitoring; and the disclosure gap management pipeline addresses reporting completeness requirements for Tadawul-listed logistics entities. The fundamental insight from this synthesis is that ESG automation is not a model selection problem but a systems architecture problem: the most viable enterprise implementation integrates a high-performing NLP layer, a robust knowledge graph, and a governance and traceability infrastructure into a coherent, auditable system.

## 10 PROPOSED METHODOLOGICAL BLUEPRINT FOR AI-ENABLED ESG REPORTING

Based on the foregoing analysis and synthesis, the following seven-step deployment blueprint is proposed as the most architecturally viable and regulatory-compliant framework for AI-enabled ESG reporting automation in logistics and supply chain contexts. The blueprint satisfies simultaneously the technical and governance requirements of ESRS, ISSB IFRS S1/S2, CSRD, and Saudi CMA ESG Disclosure Guidelines, and is structured to support phased enterprise implementation.

**Step 1:** Reporting Scope and Taxonomy Mapping. Identify all applicable reporting regimes, including CSRD/ESRS, ISSB IFRS S1/S2, EU Taxonomy, Saudi CMA ESG Disclosure Guidelines (2023), UAE SCA Sustainability Reporting Framework, GCC Exchange ESG Metrics Guidance, and sector-specific standards (GHG

Protocol, TCFD). Develop a machine-readable, version-controlled disclosure requirement graph with regulatory update triggers.

**Step 2:** Data Ingestion and Harmonisation. Build ingestion pipelines for structured metrics (emissions inventories, energy data, HR metrics) and unstructured documents (sustainability reports, supplier audit reports, contracts, policy declarations, news articles). Apply metadata standardisation, source document version tagging, data quality scoring, and provenance assignment at ingestion stage.

**Step 3:** NLP Layer. Develop and fine-tune domain-adapted transformer models for ESG text classification, entity and metric extraction, and claims identification. Prioritise ESG-specific pre-trained models [14]; for financial text, FinBERT provides a validated baseline [13]. Implement RAG architecture with disclosure content indexing for long-document processing [19]. For GCC deployments requiring Saudi CMA compliance, develop Arabic-English bilingual NLP models as a distinct and prioritised development workstream, given the absence of existing Arabic ESG domain-adapted models.

**Step 4:** Knowledge Layer. Develop a provenance-preserving ESG knowledge graph storing all extracted entities, metrics, and claims with source document links, NLP extraction timestamps, taxonomy classification tags, and human review status flags. Implement version control for taxonomy and regulatory standard updates.

**Step 5:** Predictive Analytics. Develop ESG risk and disclosure gap prediction models across all four forecast targets: ESG score prediction, financial materiality forecasting, disclosure gap prediction, and controversy/reputational event prediction. Employ ensemble and temporal models for risk prediction; implement robustness testing against label disagreement and regulatory standard drift [9, 11]. Calibration reporting is recommended for all risk score outputs deployed in thresholding or investment decision support contexts.

**Step 6:** Reporting Generation and Controls. Develop disclosure templates aligned to each applicable regulatory schema. All AI-generated narrative content must be grounded in retrieved knowledge layer evidence via the retrieve-draft-verify architecture. Embed contradiction detection, uncertainty flagging, and scenario disclaimer mechanisms. Human-in-the-loop approval is required for all material disclosure statements. Implement bilingual output generation for Saudi CMA-compliant deployments.

**Step 7:** Monitoring, Assurance, and Reporting. Implement continuous monitoring of input data distributions, taxonomy version changes, and output completeness metrics. Generate the three assurance-oriented quality metrics (evidence coverage, evidence precision, narrative faithfulness) for each reporting cycle. Maintain full audit trails for all AI processing steps.

**Step 7a:** AI Governance Framework. Define: (i) model versioning protocols and rollback procedures triggered by regulatory standard updates (annual ESRS cycles, ISSB revisions, CMA updates); (ii) human-in-the-loop approval gates for all material disclosure statements prior to publication; (iii) audit trail retention policies (minimum five years per CSRD; equivalent standards under Saudi CMA guidelines); (iv) regulatory change management procedures aligned with annual review cycles; and (v) escalation protocols for AI outputs flagged as low-confidence, low-evidence-coverage, or high-uncertainty during assurance review.

This blueprint provides a measurable, reproducible, and regulatory-compliant architecture applicable across both EU and GCC reporting jurisdictions, supporting academic reproducibility assessment and enterprise deployment governance simultaneously.

## 11 FUTURE RESEARCH DIRECTIONS

Based on the synthesis conducted in this review, six priority research directions are identified for advancing the maturity, deployability, and global applicability of AI-enabled ESG reporting automation systems:

First, the development of richer, higher-resolution public benchmarks for ESG disclosure extraction represents an urgent methodological priority. Existing benchmarks lack the granularity required by ESRS and ISSB disclosure-level annotation and provide minimal multilingual support for international supplier networks. Benchmark development specifically incorporating Arabic-language ESG corpora is a distinct gap with direct relevance to GCC and Vision 2030 deployment contexts.

Second, the ESG AI evaluation paradigm must advance substantially beyond classification metrics. Research is needed on standardised faithfulness evaluation protocols, assurance readiness measurement frameworks, provenance audit tooling standards, and human-in-the-loop evaluation protocols formally integrated with ISAE 3000 and analogous assurance standards.

Third, predictive analytics capabilities require development in calibration methodology and causal inference application. Calibration reporting should be adopted as standard practice in ESG risk modelling studies to prevent misuse of ESG risk scores as absolute truth under conditions of rating uncertainty and disagreement [9, 12]. Causal inference methods offer a technically defensible pathway beyond correlation-based ESG risk prediction for regulatory and investment applications.

Fourth, logistics operational data streams including fleet telematics, port throughput analytics, cold-chain monitoring outputs, and route optimisation data represent a significantly underexplored resource for linking operational decisions directly to ESG outcomes through predictive analytics. Research integrating operational data with ESG disclosure pipelines would substantially strengthen the managerial relevance and forward-looking precision of logistics ESG intelligence.

Fifth, the governance dimension of AI-assisted ESG reporting requires substantial dedicated research attention, including the operationalisation of due diligence requirements for AI-assisted evidence management [8], the design of AI model governance frameworks for regulated disclosure contexts, and the development of international standards for AI use in sustainability reporting assurance.

Sixth, research into ESG reporting automation tailored to GCC and emerging market contexts represents an urgent and substantially underserved priority. Specific needs include: (a) Arabic-language ESG NLP benchmark development and annotated corpus creation; (b) GCC-specific ESG taxonomy alignment tools bridging Saudi CMA Guidelines with ISSB and ESRS; (c) ESG disclosure automation frameworks for sovereign wealth fund reporting contexts (PIF, ADIA, QIA), where governance structures and disclosure obligations differ materially from listed corporate entities; and (d) applicability studies for Vision 2030-aligned programmes, including NEOM, the National Industrial Development and Logistics Programme (NIDLP), and Saudi Ports Authority sustainability mandates.

## 12 CONCLUSION

This systematic review demonstrates that AI-driven ESG intelligence has matured substantially between 2020 and 2025, transitioning from exploratory NLP experimentation toward the design of governance-grade, assurance-ready reporting infrastructure capable of supporting mandatory sustainability disclosure obligations. Transformer-based NLP particularly domain-adapted pre-training, ESG-specific fine-tuning, and retrieval-augmented generation architectures has established the technical foundation for automated ESG text classification and evidence extraction at enterprise scale. Predictive analytics serves a critical complementary role, extending ESG intelligence from retrospective compliance documentation to forward-looking risk governance, controversy anticipation, and disclosure gap remediation.

AI-driven ESG intelligence transcends report generation: it constitutes a governance infrastructure that integrates evidence management, regulatory compliance automation, and forward-looking risk intelligence into a unified disclosure architecture. The greatest operational value for logistics and supply chain organisations lies in the cohesive integration of supplier document ingestion, Scope 3 evidence management, multi-target ESG prediction, and disclosure gap management into a system that meets ESRS/CSRD, ISSB, and Saudi CMA reporting standards simultaneously and demonstrates this compliance to assurance providers through verifiable evidence coverage, precision, and faithfulness metrics.

This review contributes to the literature in three substantive ways. First, it provides a PRISMA-structured synthesis of AI-ESG reporting automation research (January 2020–February 2025) integrating NLP pipeline design, predictive analytics, and assurance readiness frameworks within a single coherent review, differentiating itself explicitly from prior related reviews [26, 27]. Second, it proposes a deployable seven-step methodological blueprint validated against ESRS, ISSB, CSRD, and Saudi CMA ESG design constraints, with explicit Arabic-language NLP requirements and Vision 2030-aligned deployment contexts the first such blueprint to incorporate GCC regulatory requirements. Third, it establishes an assurance-oriented evaluation framework introducing evidence coverage, evidence precision, and narrative faithfulness as measurable quality metrics for compliance-grade ESG AI deployments.

The primary limitations of this review are: the search scope excluded grey literature and practitioner reports, which may contain deployment evidence absent from academic databases; the qualitative synthesis method, necessitated by methodological heterogeneity, limits direct quantitative comparability; the regulatory analysis, while expanded to include GCC frameworks, remains weighted towards EU structures; and the rapid evolution of large language model capabilities after the review period may affect reported performance comparisons. These limitations define productive boundaries for the six future research directions identified in Section 11.

## REFERENCES

[1] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

[2] Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 on corporate sustainability reporting (CSRD). Official Journal of the European Union. EUR-Lex. 2022. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2464.

[3] Regulation (EU) 2020/852 of the European Parliament and of the Council of 18 June 2020 on the establishment of a framework to facilitate sustainable investment (EU Taxonomy Regulation). EUR-Lex. 2020. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32020R0852.

[4] Commission Delegated Regulation (EU) 2023/2772 of 31 July 2023 supplementing Directive 2013/34/EU as regards sustainability reporting standards (ESRS). EUR-Lex. 2023. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202302772.

[5] International Sustainability Standards Board (ISSB). IFRS S1 General Requirements for Disclosure of Sustainability-related Financial Information and IFRS S2 Climate-related Disclosures (issued June 2023). IFRS Foundation. 2023. Available: https://www.ifrs.org/issued-standards/ifrs-sustainability-disclosure-standards.

[6] U.S. Securities and Exchange Commission. SEC adopts rules to enhance and standardize climate-related disclosures for investors. Press Release 2024-31. March 2024. Available: https://www.sec.gov/news/press-release/2024-31. [Note: Rules subject to Eighth Circuit judicial stay April 2024; implementation status unresolved as of early 2026.]

[7] U.S. Securities and Exchange Commission. The Enhancement and Standardization of Climate-Related Disclosures for Investors. Final Rule, Release No. 33-11275. 2024. Available: https://www.sec.gov/rules-regulations/2024/03/33-11275.

[8] Directive (EU) 2024/1760 of the European Parliament and of the Council of 13 June 2024 on corporate sustainability due diligence (CSDDD). EUR-Lex. 2024. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024L1760.

[9] Berg F, Kölbel JF, Rigobon R. Aggregate confusion: the divergence of ESG ratings. Review of Finance. 2022;26(6):1315–1344. doi:10.1093/rof/rfac033.

[10] Christensen DM, Serafeim G, Sikochi A. Why is corporate virtue in the eye of the beholder? The case of ESG ratings. The Accounting Review. 2022;97(1):147–175. doi:10.2308/TAR-2019-0506.

[11] He Y, Shen W. ESG rating disagreement and the cost of equity financing. Journal of Financial Markets. 2025 (article in press). doi:10.1016/j.finmar.2025.100940.

[12] Bissoondoyal-Bheenick E, Brooks R, Doan T. ESG rating disagreement: implications and aggregation. Journal of Corporate Finance. 2024;84:102534. doi:10.1016/j.jcorpfin.2023.102534.

[13] Huang AH, Wang H, Yang Y. FinBERT: a large language model for extracting information from financial text. Contemporary Accounting Research. 2023;40(2):806–841. doi:10.1111/1911-3846.12832.

[14] Schimanski T, Kraus M, Webersinke N, Bingler JA, Colesanti Senni C, Leippold M. Bridging the gap in ESG measurement: using NLP to assess corporate disclosures in ESG subdomains. Journal of World Business. 2024;59(5):101554. doi:10.1016/j.jwb.2024.101554.

[15] Lee H, Kwon O, Song G, Kim Y. Automated ESG grade assessment framework using text-based techniques. Heliyon. 2024;10(16):e36294. doi:10.1016/j.heliyon.2024.e36294.

[16] Goel T, Gupta P, Singh R, et al. TCS_WITM_2022@FinSim4-ESG: augmenting BERT with linguistic and semantic features for ESG classification. Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing (FinNLP), IJCAI-ECAI. 2022;85–92. Available: https://aclanthology.org/2022.finnlp-1.10.

[17] Mashkin I, Chan Y, Lee HC, et al. HKESG at the ML-ESG-2 shared task: exploring transformer models for ESG classification. Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing (FinNLP), IJCNLP-AACL. 2023;114–120. Available: https://aclanthology.org/2023.finnlp-1.14.

[18] Saxena A, Soni D, Kumar A, Gupta V. Nested sentiment analysis for ESG impact: leveraging FinBERT with explainable AI. Mathematics. 2024;12(21):3332. doi:10.3390/math12213332.

[19] Ahmed SR, Rahman MM, Alam MJ, Islam MT. Enhancing retrieval for ESG large language models via ESG-CID: a disclosure content index for weak supervision. arXiv preprint arXiv:2501.05614. 2025. [Preprint; not peer-reviewed at time of citation. Verify publication status before journal submission.]

[20] UNCTAD Investment Policy Monitor. European Union adopts the European Sustainability Reporting Standards (ESRS). Geneva: UNCTAD. 2023. Available: https://investmentpolicy.unctad.org.

[21] IFRS Foundation. Introduction to the ISSB and the IFRS Sustainability Disclosure Standards. London: IFRS Foundation. 2023. Available: https://www.ifrs.org/groups/international-sustainability-standards-board.

[22] European Commission. EU taxonomy for sustainable activities overview and timeline. Brussels: European Commission. 2025. Available: https://finance.ec.europa.eu/sustainable-finance/tools-and-standards/eu-taxonomy-sustainable-activities_en.

[23] PwC. The European Sustainability Reporting Standards (ESRS) delegated act: overview for preparers. PwC. 2023. Available: https://www.pwc.com/gx/en/sustainability/publications/esrs-explained.html.

[24] Deloitte. Heads Up: SEC final climate disclosure rule executive summary and implementation timeline. Deloitte Center for Financial Services. 2024. Available: https://dart.deloitte.com/USDART/home/publications/deloitte/heads-up/2024/sec-climate-related-disclosure-final-rule.

[25] Pastor L, Stambaugh RF, Taylor LA. Sustainable investing in equilibrium. Journal of Financial Economics. 2021;142(2):550–571. doi:10.1016/j.jfineco.2020.12.011.

[26] Meng Q, Zhou H, Li M, Chen J. Comparative analysis of corporate sustainability reports using NLP approaches. Sustainability. 2025;17(22):10315. doi:10.3390/su172210315.

[27] Elhady AM, Shohieb S. AI-driven sustainable finance: computational tools, ESG metrics, and global implementation. Future Business Journal. 2025;11:209. doi:10.1186/s43093-025-00610-x.

[28] EUR-Lex. CSRD Directive full text (CELEX:32022L2464). Official Journal of the European Union. 2022. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2464.

[29] EUR-Lex. EU Taxonomy Regulation full text (CELEX:32020R0852). Official Journal of the European Union. 2020. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32020R0852.

[30] Vancea DPC, Petria N, Iuga IC, Boitor B. Artificial intelligence in sustainable finance: a bibliometric analysis and systematic literature review. Sustainability Analytics and Modeling. 2024;4:100031. doi:10.1016/j.samod.2024.100031.

[31] CDP. Global Supply Chain Report 2020: Transparency to Transformation. London: Carbon Disclosure Project. 2020. Available: https://www.cdp.net/en/research/global-reports/global-supply-chain-report-2020.

[32] Greenhouse Gas Protocol. Corporate Value Chain (Scope 3) Accounting and Reporting Standard. Washington, DC: World Resources Institute and World

Business Council for Sustainable Development. 2011. Available: https://ghgprotocol.org/scope-3-standard.

[33] KPMG International. Survey of Sustainability Reporting 2022. Amsterdam: KPMG. 2022. Available: https://kpmg.com/xx/en/home/insights/2022/09/survey-of-sustainability-reporting-2022.html.

[A] Saudi Capital Market Authority (CMA). ESG Disclosure Guidelines for Listed Companies. CMA, Riyadh, Kingdom of Saudi Arabia. 2023. Available: https://cma.org.sa/en/Market/ESG/Pages/default.aspx.

[B] Saudi Exchange (Tadawul). Saudi ESG Index (SASESG): Methodology and Eligibility Criteria. Riyadh: Saudi Exchange. 2023. Available: https://www.saudiexchange.sa/wps/portal/saudiexchange/index-management/sasesg.

[C] Public Investment Fund (PIF). PIF Sustainability Report 2023. Riyadh: PIF. 2023. Available: https://www.pif.gov.sa/en/Pages/sustainability.aspx.

[D] Kingdom of Saudi Arabia. Saudi Green Initiative: Framework Document. Saudi Green Initiative Secretariat. 2021. Available: https://www.saudigreeninitiative.org.

[E] UAE Securities and Commodities Authority (SCA). Sustainability Reporting Guidelines for Listed Companies. Abu Dhabi: SCA. 2023. Available: https://www.sca.gov.ae/en/pages/sustainability.aspx.

[F] GCC Exchanges Committee. ESG Reporting Guidance for GCC-Listed Companies. Manama: GCC Exchanges Committee. 2022. Available: https://www.gccexchanges.org/esg-reporting-guidance.

## Authors' Contribution

All authors contributed equally to the development of this article.

## Data availability

All datasets relevant to this study's findings are fully available within the article.

## How to cite this article (APA)

Qureshi, O. U. R. (2026). AI-ENABLED ESG INTELLIGENCE: A SYSTEMATIC REVIEW OF NLP AND PREDICTIVE ANALYTICS FRAMEWORKS FOR AUTOMATED SUSTAINABILITY REPORTING (2020–2025). *Veredas Do Direito*, 23(5), e235225. https://doi.org/10.18623/rvd.v23.5225