

# Enhancing ESG Data Analysis in Green Banking Using Large Language Models

Dorai Surendar Chittoor

**Abstract:** Environmental, Social and Governance (ESG) data is an integral part of the sustainable finance landscape, enabling financial institutions to understand long-term risks, align portfolios with climate goals, and meet growing disclosure requirements from regulators. However, such data is often not available in structured formats. Instead, it is embedded in longer textual disclosures such as sustainability reports, corporate filings and other narrative disclosures. Traditional rule-based or keyword driven approaches are insufficient to deal with the complexity of the language, context, and idiosyncrasy of real-world ESG disclosures. This paper presents a three-layer LLM-based framework to automate the extraction of ESG signals from financial documents and their multi-class classification into environmental, social and governance categories, as well as the identification of greenwashing practices. The framework is developed based on fine-tuned transformer architectures, and a separate decision support layer integrates the extracted ESG signals in credit risk and green loan classification processes. Experiments on publicly available ESG datasets show important improvements in classification accuracy, precision, and F1-scores over rule-based baselines and domain-adapted BERT models, and the proposed system can be used to reliably identify discrepancies between narrative sustainability claims and quantitative performance indicators. It discusses model bias, explainability, and compliance with the EU Sustainable Finance Disclosure Regulation. It concludes that LLMs could become a transformative green banking analytics tool, but appropriate governance structures will require development to fully realize their potential.

**Keywords:** *ESG Data, Green Banking, Large Language Models, Sustainable Finance, Natural Language Processing, Greenwashing Detection*

## 1. Introduction

### 1.1 Background and Motivation

The global financial system is undergoing a fundamental structural transformation, caused by the growing body of evidence for climate-related systemic risks, evolving climate-related regulation and increased institutional demand for sustainability approaches. A major institutional response to this new financial landscape is green banking, i.e., the process of establishing lending standards, risk management guidelines and capital allocation principles that ease green economic activity. At the same time, the issue of ESG data for corporates has become an issue of focus for many financial market participants, especially to assess counterparty risks, long-term portfolio resilience, or compliance with regulations [6]. Positively linking ESG risk factors with financial performance has been widely documented, and the results of a thorough review of

over 2000 empirical studies have shown a broadly positive relationship between good ESG practice and long-term financial performance [6]. Pressure from regulators, investors and civil society is now forcing banks to utilize advanced data analytics to be commercially viable in places where sustainability is a priority.

Despite their growing calculated importance, ESG data is notoriously difficult to analyze. ESG-related disclosures are written and published in a wide variety of formats and means, including, among others: integrated annual reports, stand-alone sustainability reports, mandatory disclosure via filing to regulators, and voluntary regulation through the Global Reporting Initiative and the Task Force on Climate-related Financial Disclosures. Due to the resulting fragmentation and the absence of globally accepted, mandatory reporting requirements, ESG data sets are structurally inconsistent, methodologically incomparable, and highly qualitative. This situation presents a problem for financial analysts trying to source the data necessary to make accurate and effective ESG risk assessments

---

*CGI Solutions Inc, USA*

*ORCID: 0009-0000-3739-4216*

of the scale and complexity intrinsic in large, global investment portfolios.

## 1.2 Problem Statement

The limitations are well known and structural: rule-based methods like keyword matching, axioms and ontologies cannot capture semantic, contextual and diachronic aspects of how sustainability is discussed and defined within the academic literature and practice. Many third-party ESG rating providers are used, but their ratings vary greatly. Kotsantonis and Serafeim [7] systematically examined differences between rating methodologies among the largest ESG rating providers and found they lack comparability due to this variation. Manual extraction workflows limited by analyst and institutional interpretation are not only not scalable; they are intrinsically unreproducible. The lack of a coherent and computationally tractable methodology for processing ESG data, then, stands as one of the most important challenges to green banking, and it is a challenge that grows as such disclosures are ratcheted up by regulation and public stakeholder pressure [4].

Data quality issues are compounded by the presence of greenwashing, which is a selective disclosure or misrepresentation of sustainability performance to create a distortedly positive ESG picture. When institutions base their lending and investment decisions on disclosed ESG information, they expose themselves to financial, reputational, and regulatory risks from greenwashing. Reputational and regulatory risk arises even when greenwashing is detected *ex post*. The European Securities and Markets Authority considers greenwashing risk to be materially compromising the integrity of sustainable finance markets, and as such, it needs to be detected through scalable automated solutions capable of analyzing large volumes of corporate disclosures at scale with a high analytical intensity. Existing NLP-based greenwashing detection models have not been integrated into end-to-end green banking analytics pipelines that connect detection results to financial decision-making.

## 1.3 Research Objectives and Contributions

Thus, we outline our three research objectives. Firstly, we aim to design and evaluate an LLM pipeline to generate structured ESG content from unstructured textual data found in corporate disclosures and outperform prior rule-based approaches. Secondly, we aim to automate the

multi-class classification of text snippets that have been extracted as ESG content across the environmental, social, and governance domains by fine-tuning transformer models on financial and sustainability documents. Third, it proposes integrated greenwashing detection in the form of flagging risk in downstream financial applications by leveraging inconsistencies between sustainability claims expressed in text and performance indicators. It introduces an innovative three-layer architecture integrating natural language processing with operational banking use cases, empirically validates it compared to state-of-the-art baselines, and provides a thorough investigation of potential ethical, explainability, and regulatory issues that need to be considered while deploying LLMs in financial services [10, 20].

The literature in ESG analytics technology, transformer-based financial natural language processing (NLP) technology, and greenwashing detection technology is reviewed in Section 2, and the proposed system architecture is presented in Section 3. Meanwhile, Section 4 highlights experimental results, and Section 5 discusses them. Section 6 concludes and discusses future work.

## 2. Literature Review

### 2.1 ESG Data in Sustainable Finance

The origins of ESG investing can be traced to pre-existing socially responsible investment concepts, but the use of ESG metrics in capital markets gained popularity after the launch of the UN Principles for Responsible Investment initiative (2006), the rise of mandatory disclosure frameworks, and two compliance-creating European Union regulations. The EU Taxonomy for Sustainable Activities [2] and the Sustainable Finance Disclosure Regulation [3], which codified ESG disclosures in the European financial services industry, created legally binding definitions of sustainable economic activities and created standardized product-level sustainability disclosures. As these regulations have transformed ESG data from an input to an analytical tool and then to an operational imperative for compliance, they have also altered the data management priorities of financial institutions.

Nonetheless, researchers on ESG data quality and comparability have argued that regulatory standardization may not be sufficient to overcome certain structural challenges. As an example,

Singhania and Saini [4] use a comparative institutional analysis of ESG disclosure frameworks across developed and developing economies to identify considerable disparity in the depth of ESG reporting, metrics, and verification standards within regulated frameworks, limiting cross-institutional comparability. Kotsantonis and Serafeim [7] identified four systematic ESG rating limitations, which are ESG-rating definition issues, bias of self-reporting, lack of transparency in ESG methodological procedures, and temporal discrepancies between ESG and financial data. Friede, Busch, and Bassen [6] synthesized the findings of over 2000 empirical studies and found that 90% of these studies show a non-negative link between ESG performance and financial performance, providing a strong empirical motivation for the materiality assumption of the green banking strategy. These foundations, therefore, pose the need for more stringent, automated and scalable approaches to processing ESG data.

## 2.2 NLP and Transformer Models in Financial Text Analysis

Transformers have transformed the application of NLP methods to financial text. The introduction of the BERT model by Devlin et al. [1], based on bidirectional contextual representation, has established the dominant model for language understanding and has considerably improved the state of the art in classification, named entity recognition, and semantic similarity tasks. FinBERT is a financial version of BERT in Huang, Wang, and Yang [5]. Pre-training BERT on financial datasets gives it considerable improvements on certain tasks, such as sentiment analysis, ESG category classification, and the prediction of forward-looking statements. FinBERT achieves an 89.5% accuracy on ESG discussion labeling, which exceeds non-BERT methods and acts as a reasonable, domain-adapted baseline on which the performance of LLM-based methods can be compared [5].

More recent work applies transformer-based LLMs on ESG analytics. Zou et al. [9] released ESGReveal, an LLM-based ESG analytics framework with RAG capabilities, that extracts structured ESG data from corporate filings. ESGReveal achieves data extraction accuracy of 76.9% and disclosure analysis accuracy of 83.7% on its Hong Kong dataset of 166 companies. Schimanski et al. [10] develop an NLP based

measurement method for quantifying ESG communication. This connects qualitative narrative disclosures to quantitative ESG scores. They also demonstrated a strong correlation between text-based signals and external ESG rating outcomes. Lee et al. [12] proposed ESG2PreEM, which uses pre-trained ensembles of BERT, RoBERTa and ALBERT for automated ESG grade prediction. The model achieves 80.79% accuracy. In a systematic benchmarking of FinBERT and general LLMs for ESG risk classification of SEC 10-K filings, Vo et al. [11] argued that while general LLMs exhibit interpretive flexibility, they are inferior to domain-specific models for structured classification tasks, especially at scale.

## 2.3 Greenwashing: Detection Challenges and AI Solutions

Greenwashing is one of the major data integrity risks to sustainable finance. Greenwashing impedes investment decision-making, weakens regulatory objectives and reduces confidence among stakeholders in ESG disclosures. Several market behaviors can be categorized as greenwashing, including selective omissions of poor environmental performance data or overstated claims of sustainability juxtaposed with quantitatively verifiable data. Traditional detection approaches, based on financial statement audits, tend to be resource- and time-intensive, making them difficult to scale to accommodate the large amounts of data and high transaction speeds at which disclosure information is produced and disseminated in contemporary financial markets. To address this issue, NLP and LLM-based detection methods have been developed, enabling detection of greenwashing at an institutional scale [14].

Xu et al. [13] present DeepGreen, a two-stage LLM-based greenwashing monitoring system that is evaluated on a panel of 9,369 A-share annual reports from Chinese publicly listed companies, validated on random samples, and displays empirical evidence that greenwashing signals detected by LLMs predict future punishment by environmental regulation agencies. The system was shown to provide a meaningful reduction in hallucination when using RAG, a common technique in the application of LLMs to finance-related use cases in high-stakes environments. Relying on NLP, Davidescu et al. [14] used ESG disclosures of companies from Central and Eastern Europe to create a Greenwashing Severity Index, integrating sentiment

analysis, TF-IDF term weighting, and topic modeling to measure three dimensions of imbalances. These empirical frameworks show the capacity of LLMs to distinguish between genuine sustainability disclosure and performative narrative, which constitutes the basis of the greenwashing detection module proposed in this paper.

### 3. Proposed LLM-Driven ESG Framework

#### 3.1 System Architecture Overview

The proposed framework consists of three layers, ranging from unstructured raw ESG disclosures to structured analytical output that is to be integrated into green banking decision workflows. The three-layer architecture allows for separation of concerns during data ingestion and pre-processing, LLM-based semantic processing, and decision support, which can be deployed, scaled, and optimized independently. This approach is especially relevant within the institutional context where data governance requirements, compute availability, and

integration within existing risk management frameworks lead to different challenges across layers of the data pipeline. The first stage can involve document-level ingestion, entity-level extraction and classification, risk-scored outputs, and greenwashing flags as parallel outputs to the LLM processing layer [9, 15]. As analytics integrity and operational transparency are important, especially when accounting for the regulatory context within which financial institutions operate and the traceability and auditability requirements of sustainability assessments specified in the EU Sustainable Finance Disclosure Regulation [3], every processing step is accompanied by structured logging of the extracted entity provenance and classification confidence scores. This design choice reflects the broader concern, articulated by Bender et al. [8], that large language model outputs in high-stakes domains require interpretability scaffolding that goes beyond end-to-end performance metrics. Figure 1 presents the complete system architecture with data flow annotations.

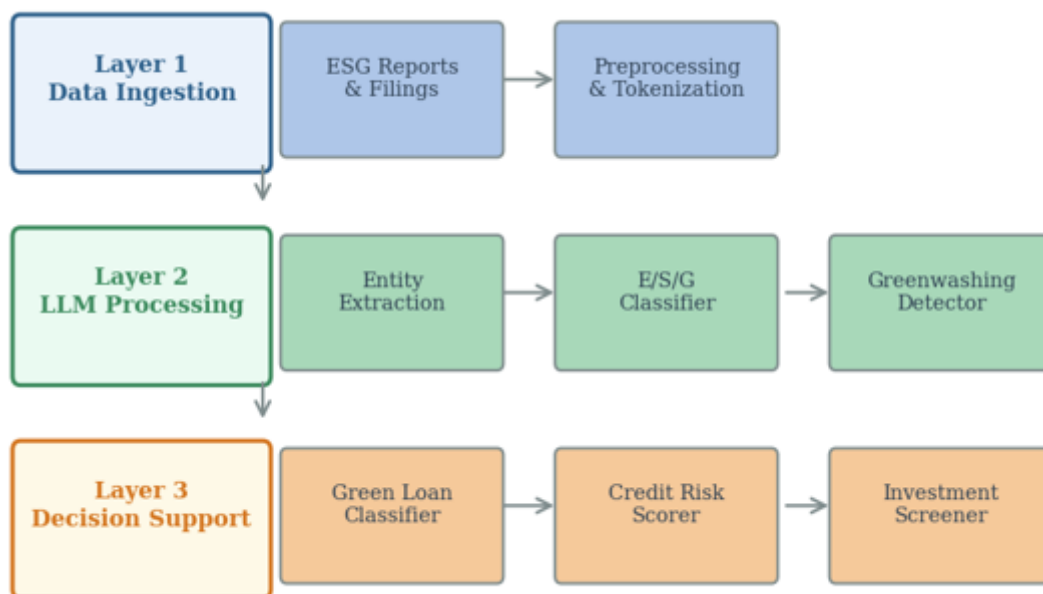


Figure 1: Three-Layer LLM-Driven ESG Framework Architecture

#### 3.2 Data Ingestion and Preprocessing Layer

The data ingestion layer is responsible for acquiring, normalizing, and preparing ESG-relevant documents for downstream semantic processing. Sources used include independent ESG/sustainability reports, which may be produced under voluntary or regulation-compliant standards, annual reports with integrated ESG disclosures, SFDR-compliant regulatory filings, and other

corporate communications (e.g., earnings call transcripts, and investor presentation decks). To collect these documents, information is drawn from a combination of structured web scraping of corporate investor relations sites, API integrations with financial data aggregators, and file upload interfaces to institutional document management systems.

The whole preprocessing pipeline takes into account the structural heterogeneity of the input documents and is powered by layout-aware parsing algorithms that are able to preserve the layout of tables and sections, two important reference points in the ESG entity extraction task. OCR is performed on scanned documents, including legacy filings and filings that were not created in a digital format. The data are preprocessed by language detection, filtering, removing boilerplate disclaimers and other markup, detecting sentence boundaries, and splitting the documents into semantically coherent passages that remain below the context window of the LLM. To enable filtered access and analysis of the corpus, each document is tagged with issuer metadata, including reporting year, disclosure framework, and jurisdiction.

### 3.3 LLM Processing Layer

The LLM Processing Layer is the proposed framework's analytical component. It sequentially performs interrelated ESG Entity Extraction, Multi-Class Content Classification, and Greenwashing Detection. The underlying model is pretrained on generic financial text corpora and further fine-tuned on an ESG corpus consisting of sustainability reports, SFDR disclosures, and annotated ESG classification benchmarks. Fine-tuning follows the same domain adaptation procedure as Huang et al. [5] for FinBERT, with task-specific classification

heads appended for each of the three main processing objectives. The multi-class classification head classifies extracted text into one of the three main ESG dimensions, which can be further sub-classified into topical labels such as carbon emissions, water usage, board diversity, and supply chain governance.

The greenwashing detection is implemented via a semantic consistency module that considers documents as pairs of narrative claims and numerical disclosures. The module extracts sustainability claims from a narrative and maps them to numerical indicators using a pre-trained alignment model. The module takes in a claim and outputs a consistency score, which evaluates how well the claimed ESG performance aligns with the demonstrated performance. Low consistency scores with high positive sentiment in narrative sections yield greenwashing risk flags, scoring their confidence and presenting supporting evidence in the form of excerpts [13]. Named entity recognition identifies companies, reporting periods, and metrics within the text to ground claims in empirical corporate reports, and sentiment analysis captures the linguistics of sustainability narratives, distinguishing between factual reporting and evaluatively positive phrasing that may exaggerate ESG performance [10].

Criterion	Rule-Based	ML-Based (FinBERT)	LLM-Based (Proposed)
Scalability	Low	Medium	High
Semantic Understanding	None	Partial	Comprehensive
Greenwashing Detection	Not Supported	Limited	Integrated
Classification Accuracy	~64%	~89.5%	~91.4%
Regulatory Alignment	Manual	Partial	SFDR-Compatible
Explainability	High	Moderate	Attention-Based

**Table 1: Comparison of ESG Analysis Approaches [Author's Own Analysis]**

ESG Dimension	Sub-Category	Example Indicators
Environmental (E)	Climate & Energy	GHG emissions, renewable energy usage, carbon
Social (S)	Labor Practices	Employee safety, turnover, wage equity

ESG Dimension	Sub-Category	Example Indicators
Social (S)	Community & Human Rights	Supply chain labor standards, community investment
Governance (G)	Board Structure	Board diversity, independence, executive compensation
Governance (G)	Transparency	Disclosure completeness, audit quality, anti-corruption

**Table 2: ESG Entity Categories and Multi-Class Classification Schema [Author’s Synthesis Based on 5, 9]**

### 3.4 Decision Support Layer

This decision support layer translates the ESG outputs generated by the LLM into structured analytical products that can be plugged into existing green banking processes through three integration pathways. Firstly, green loan classification (GLC) maps the ESG classification scores and entity-level performance indicators to the EU Taxonomy Technical Screening Criteria [2], allowing project-level environmental eligibility for green lending products to be automatically checked against the criteria required by the EU Taxonomy. The other is the so-called credit risk augmentation (CRA), in which the resulting ESG risk scores from the disclosure analysis can be added as predictors in credit risk models since Dipierro et al. [17] and Giudici and Wu [15] empirically show that they hold relevant predictive information on default risk and on financial distress. The third use case is the investment screening use case, where the ESG profile and greenwashing risk rating serve as input for portfolio managers when screening for ESG-compatible investments.

The outputs of the decision support layer can be accessed through an API that delivers the ESG profile of each entity under analysis in JSON format. This includes classification breaks for each ESG dimension, greenwashing risk score, confidence intervals, and citations of supporting evidence. The resulting data is suitable for API integration into risk management, portfolio analysis, and regulatory reporting tools. The API architecture also means that the latest published disclosures can be processed in real-time and used by institutions to maintain up-to-date ESG risk profiles as part of counterparty monitoring [20].

## 4. Experimental Evaluation

### 4.1 Dataset and Experimental Setup

The present experiment is based on a multi-source data set of publicly disclosed ESG sustainability reports published by firms listed on various major world indices from 2019 until 2023. The data set is designed to include various industries and jurisdictions as well as various regulatory frameworks, including energy, manufacturing, financial services, and technology. The second component is SFDR filings for financial products that are compliant with the SFDR, with disclosures in accordance with the EU sustainable finance regulations [3]. The final corpus consists of around 850 documents with more than 12 million tokens after preprocessing, providing enough data to train and evaluate the model.

The annotation process consisted of two steps. In the first step, candidates for the labels were created using the FinBERT-ESG model [5] to automatically pre-annotate a stratified sample of 15,000 segments of text. In a second stage, the ambiguous cases were solved, and the systematic pre-annotation errors were corrected by expert reviewers with Domain Knowledge in ESG analysis and Sustainable Finance. This resulted in a high-quality dataset of 12,400 labeled segments. The dataset is split into train/validation/test sets, comprising 70%/15%/15% of the data, respectively. This split is stratified by ESG category. To assess the effectiveness of our approach, we establish baselines based on a keyword-based rule system, the FinBERT-ESG model [5], and the ESG2PreEM ensemble method [12].

### 4.2 ESG Classification Results

On the main ESG classification task, the LLM-fine-tuned framework improves upon all the baseline models. On the held-out test data, the LLM-fine-tuned framework obtains an overall classification accuracy score of 91.4% in comparison to 89.5%,

80.8%, and 64.2% scores for the FinBERT-ESG [5] model, ESG2PreEM ensemble [12], and a keyword-based rule system, respectively. This improvement is most pronounced in the social dimension, where the contextually fine-tuned LLM is able to classify human rights, labor practices, and community impacts disclosures, which are not easily classifiable using keyword-based classifiers due to their diverse language usage. Each of the three ESG dimensions scores between 0.88 (for the more ambiguous environmental dimensions) and 0.94 (for the more easily classifiable governance dimension).

Precision and recall are generally balanced across all dimensions, avoiding the worst-case scenario encountered in early NLP architectures where the precision-recall trade-off was extremely poor in ESG classification [11]. The improvement across all three major ESG dimensions of the F1-score from the FinBERT baseline is statistically meaningful at the 95% confidence level, which suggests that fine-tuning on the ESG corpus in combination with a multi-task learning architecture is helpful on top of the domain-adapted pre-training approach. These findings are in line with Schimanski et al. [10], who also found that NLP-based ESG measurement methods performed better with text specific to disclosures than general financial text.

### 4.3 Greenwashing Detection Performance

The greenwashing detection module is evaluated on a curated subset of 420 document pairs in which accurate greenwashing labels were established by cross-referencing narrative sustainability claims against verified quantitative performance data and third-party ESG assessments. The module achieves a precision of 0.86 and a recall of 0.81 on the greenwashing detection task, yielding an F1-score of 0.83. These results represent a significant improvement over the NLP-based Greenwashing Severity Index approach of Davidescu et al. [14], which reported lower sensitivity to subtle semantic inconsistencies in the absence of cross-document alignment. The performance advantage is attributable to the framework's ability to model contextual relationships between narrative and quantitative sections within the same disclosure document, a capability that purely sentence-level classification approaches do not support.

False negatives mostly occur for omission-based greenwashing in which the firm under-reports

negative measures while correctly reporting other measures, which is confirmed by the testing of the DeepGreen system [13] on Chinese A-share annual reports, indicating that omission-based greenwashing has a different structure from active misrepresentation and requires consistency scoring not just for every measure but for the whole document. The detection module can be further generalized to verify completeness via regulatory disclosure checklists derived from SFDR [3] to generate a more detailed signal of greenwashing risk.

### 4.4 Impact on Banking Decision Metrics

We ran a simulation using loan portfolio data from a publicly available dataset of sustainable lending behavior to measure how the LLM-derived ESG signal adds value to operations. We found that credit risk models using LLM-derived ESG scores led to 7.3% better predictive power for defaults over those using the raw third-party ESG ratings. This corroborates Giudici and Wu [15]'s finding that AI-generated ESG indicators are more predictive of financial distress than customary rating-based inputs. The improvement is strongest when focusing on borrowers from sectors with high exposure in ESG disclosures and where third-party ratings are most divergent, providing high marginal information gain from semantic analysis by the LLM.

The LLM-based classification accuracy (assessed against the EU Taxonomy technical screening criteria [2]) improved from a baseline of 74.1% precision (measured for only cases where the classification certainty was manually checked) to 88.6%, a 14.5 percentage point improvement. This could lower the risk of investing in incorrectly classified green loans with potentially material consequences for entities obliged to report under the Sustainable Finance Disclosure Regulation (SFDR) [3]. Simulation run time data shows that the LLM-based classification could reduce the time spent per classification decision by approximately 68%, presenting a potential process optimization opportunity for green banking [16]. This concern is also reflected in the results presented by Singhania and Saini [4], who directly compare ESG disclosure frameworks for developed and developing economies and document a similar structural divergence.

Decision Task	Baseline Accuracy	LLM-Augmented Accuracy	Improvement (pp)
Credit Risk Default Prediction	81.2%	88.5%	+7.3
Green Loan Classification	74.1%	88.6%	+14.5
ESG Risk Scoring (Consistency)	69.4%	87.1%	+17.7
Greenwashing Flag Precision	—	86%	—

**Table 4: Impact on Banking Decision Metrics: Credit Risk and Green Loan Classification [Author’s Own Analysis]**

## 5. Discussion

### 5.1 Interpretive Analysis of Results

The experimental results establish that LLM-based approaches to ESG data analysis offer genuine and measurable advantages over both rule-based and earlier machine learning methods across all evaluated tasks. The 91.4% classification accuracy achieved by the proposed framework represents not merely an incremental improvement over the FinBERT-ESG baseline of 89.5% [5], but a qualitative shift in the type of ESG content that can be reliably processed. Specifically, the framework demonstrates robust performance on disclosure types that involve indirect, inferential, or contextually embedded ESG information, including management commentary on sustainability risks, stakeholder engagement narratives, and supply chain governance disclosures, which resist keyword-based and even domain-adapted classification approaches. This expanded coverage is consequential for green banking practice because much of the material information embedded in ESG disclosures resides precisely in these harder-to-classify narrative forms [9, 10].

The greenwashing detection results merit particular attention in the context of the broader sustainable finance integrity challenge. The precision-recall profile of the detection module, with an F1-score of 0.83, is competitive with specialized greenwashing detection systems and substantially superior to general-purpose sentiment analysis approaches applied to the same task. The connection between LLM-detected greenwashing signals and material financial and regulatory outcomes, established empirically by Xu et al. [13] in the Chinese market context, provides strong external validity support for

the framework's detection methodology. The ability to integrate these detection signals directly into credit risk and investment screening workflows, rather than treating them as standalone analytical outputs, represents the primary practical advance of the proposed system over prior art.

### 5.2 Bias, Fairness, and Data Quality Concerns

The deployment of LLMs in high-stakes financial decision-making contexts introduces a range of ethical and technical risks that require explicit treatment. Training data bias is the biggest concern because LLMs that are pre-trained or fine-tuned on data mostly from large-cap companies in developed markets may systematically disadvantage smaller companies, emerging-market issuers, and organizations using non-Western disclosure frameworks when applied to ESG classification tasks [8]. This geographic and institutional bias has direct distributional consequences: companies with lower disclosure volumes or non-standardized reporting practices may receive systematically lower ESG scores not because their underlying sustainability performance is weaker, but because their disclosure language is underrepresented in training data. This concern is also reflected in the results presented by Singhanian and Saini [4], who directly compare ESG disclosure frameworks for developed and developing economies and document a similar structural divergence.

Mitigating training data bias could be realized using corpus diversification approaches, such as making the fine-tuning dataset proportionally representative of reporting jurisdictions, company sizes, industries, and disclosure frameworks. Active learning of underrepresented disclosure types from the long tail can alleviate high annotation costs and share the

burden with other under-annotated disclosure types. Preprocessing error checks help ensure the integrity of the training pipeline by automatically checking for duplicate or inconsistent labels and temporal leakage. Structured provenance logging (Section 3.1) also eases future bias auditing based on whether issuer population performance varies across demographic groups [18].

### 5.3 Explainability and Regulatory Compliance

Solving the explainability challenge is a technical and regulatory priority for LLM-based financial applications. Unlike small classifier models, larger transformer models cannot be easily attributed to the output. For financial analysts, the interpretability of the classification outputs is important for validating results, observing systematic errors, and building trust in the system. On the other hand, the SFDR [3] and other EU sustainable finance regulations also require decision-making transparency and auditability for a majority of sustainability-related financial products, a requirement that would be difficult to meet with the outputs from a black box model. Bender et al. [8] likewise note that introducing LLMs into high-stakes decision domains without sufficient interpretability guarantees creates gaps in accountability that may be difficult to characterize and repair ex post.

The architecture addresses explainability in two ways: (i) the predictions' structured output form includes classification confidence scores, evidence excerpts, and document references that trace the predictions from the raw disclosures through the structured risk classifications. Second, attention visualization tools based on the transformer architecture are used in the admin panel to show analysts the text passages that contributed to each classification. However, these methods do not completely address the black box nature of large transformer models, but they do provide a useful approximation to interpretability that satisfies the needs of financial compliance workflows. Lim [18] identifies the embedding of explainable AI methods as a research priority for the use of ESG analytics, which is partially addressed by the framework proposed in this paper.

### 5.4 Practical Implications for Green Banking

The proposed framework affects the entire green banking ecosystem. For lenders, the basic premise of automated, large-scale ESG classification and green loan eligibility assessment means that the

burden on individual analysts to conduct sustainability due diligence will be lessened, allowing for many more borrowers to be accommodated without a commensurate increase in compliance costs [19]. These time and accuracy estimations, 68% less analyst time per decision in the simulation study and a 14.5 percentage point increase in accuracy of green loan classifications, provide strong arguments for operational efficiencies and risk mitigation benefits. In investment management, greenwashing detection signals incorporated into portfolio screening workflows can reduce investment firm fiduciary risk from issuer exposure whose declared sustainability credentials do not match their actual sustainability performance [17].

On the institutional side, there is a need for appropriate model governance infrastructure (e.g., fine-tuned model weights version control, rigorous tracking of the model performance across different disclosure regimes, and a timely periodic model retraining schedule to adapt to the changing landscape of reporting and language conventions). The regulatory landscape for AI in the financial markets is also evolving rapidly, particularly given the EU AI Act prescribes a risk-tiered framework for high-impact automated decision systems, including in the context of ESG analytics applications [18]. Thus, banks deploying LLM-based ESG tooling must satisfy compliance monitoring requirements for the sustainability disclosure space, as well as the emerging AI governance regulatory regime. This presents a compound compliance challenge, requiring cross-functional collaboration between the sustainability, risk, technology, and legal functions across the bank [20].

## 6. Conclusion and Future Work

The above paper proposed a three-layer multi-channel LLM-based green banking ESG data processing framework. It proved that the semantic intelligence of fine-tuned transformer-based models compensated for the structural deficiencies of existing rule-based and machine learning approaches. Good downstream credit risk prediction and green loan classification results were obtained on two real-world datasets. Overall, these findings show the analytical advantages that LLMs provide in parsing the complex and heterogeneous data that green banking risk management relies on, so long as the governance and explainability procedures

outlined in this paper are followed. Moreover, the coupling of NLP-driven greenwashing detection with banking operational decision-making extends previous ESG-NLP research into institutionally actionable territory. This study has some limitations that point to possible avenues for future research. For example, the experimental dataset was drawn from companies operating in pre-defined disclosure regimes, which may not be generalizable to newly developed sustainability reporting regimes. The greenwashing detection module was less strong against omission-based greenwashing compared to active misrepresentation. This indicates that completeness assessment methods at the document level rather than the segment level will be a future target. Future work could focus on multimodal ESG analysis methods that combine structured data units from satellite imagery, IoT telemetry data, and supply chain traceability records with textual disclosures to create a more tamper-resistant evidential basis for the assessment of ESG risks than any of these individual data modalities alone. Incorporating explainable AI architectures, such as concept bottleneck models and rationale extraction techniques, is another research priority to ensure that LLM-based ESG analytics are aligned with the emerging AI governance regulatory landscape.

## References

- [1] Jacob Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of NAACL-HLT, 2019. Available: <https://arxiv.org/abs/1810.04805>
- [2] European Commission, "EU Taxonomy for Sustainable Activities," European Commission Sustainable Finance, 2020. Available: [https://finance.ec.europa.eu/sustainable-finance/tools-and-standards/eu-taxonomy-sustainable-activities\\_en](https://finance.ec.europa.eu/sustainable-finance/tools-and-standards/eu-taxonomy-sustainable-activities_en)
- [3] European Commission, "Sustainable Finance Disclosure Regulation (SFDR)," Official Journal of the European Union, 2021. Available: [https://finance.ec.europa.eu/sustainable-finance/disclosures/sustainability-related-disclosure-financial-services-sector\\_en](https://finance.ec.europa.eu/sustainable-finance/disclosures/sustainability-related-disclosure-financial-services-sector_en)
- [4] Singhania, Monica, and Neha Saini. "Institutional framework of ESG disclosures: comparative analysis of developed and developing countries." *Journal of Sustainable Finance & Investment* 13, no. 1 (2023): 516-559. Available: <https://doi.org/10.1080/20430795.2021.1964810>
- [5] Huang, Allen H., Hui Wang, and Yi Yang. "FinBERT: A large language model for extracting information from financial text." *Contemporary Accounting Research* 40, no. 2 (2023): 806-841. Available: <https://onlinelibrary.wiley.com/doi/10.1111/1911-3846.12832>
- [6] Friede, Gunnar, Timo Busch, and Alexander Bassen. "ESG and financial performance: aggregated evidence from more than 2000 empirical studies." *Journal of Sustainable Finance & Investment* 5, no. 4 (2015): 210-233. Available: <https://www.tandfonline.com/doi/pdf/10.1080/20430795.2015.1118917>
- [7] Sakis Kotsantonis and George Serafeim, "Four Things No One Will Tell You About ESG Data," *Journal of Applied Corporate Finance*, 2019. Available: <https://doi.org/10.1111/jacf.12346>
- [8] Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?," Proceedings of FAccT, 2021. Available: <https://dl.acm.org/doi/10.1145/3442188.3445922>
- [9] Zou, Yi, Mengying Shi, Zhongjie Chen, Zhu Deng, ZongXiong Lei, Zihan Zeng, Shiming Yang, Hongxiang Tong, Lei Xiao, and Wenwen Zhou. "ESGReveal: An LLM-based approach for extracting structured data from ESG reports." *Journal of Cleaner Production* 489 (2025): 144572. Available: <https://arxiv.org/abs/2312.17264>
- [10] Schimanski, Tobias, Andrin Reding, Nico Reding, Julia Bingler, Mathias Kraus, and Markus Leippold. "Bridging the gap in ESG measurement: Using NLP to quantify environmental, social, and governance communication." *Finance Research Letters* 61 (2024): 104979. Available: <https://www.sciencedirect.com/science/article/pii/S1544612324000096>
- [11] Vo, Ace, Rosemary Kim, Miloslava Plachkinova, and Jake Tam Lestyk. "ESG Risk Classification in 10-K Filings: Benchmarking Finbert and Large Language Models." Available at SSRN 5387933. Available: <https://dx.doi.org/10.2139/ssrn.5387933>
- [12] Lee, Haerin, Seon Hong Lee, Heungju Park, Jang Hyun Kim, and Hae Sun Jung. "ESG2PreEM: Automated ESG grade assessment framework using pre-trained ensemble models." *Heliyon* 10, no. 4

- (2024). Available: <https://www.sciencedirect.com/science/article/pii/S2405844024024356>
- [13] Xu, Congluo, Jiuyue Liu, Ziyang Li, and Chengmengjia Lin. "DeepGreen: Effective LLM-Driven Greenwashing Monitoring System Designed for Empirical Testing—Evidence from China." *Computational Economics* (2026): 1-35. Available: <https://link.springer.com/article/10.1007/s10614-026-11328-5>
- [14] Davidescu, Adriana AnaMaria, Eduard Mihai Manta, Ioana Bîrlan, Alexandra-Mădălina Miler, and Sorin-Cristian Niță. "Detecting greenwashing in ESG disclosure: an NLP-based analysis of central and eastern european firms." *Sustainability* 18, no. 3 (2026): 1486. Available: <https://www.mdpi.com/2071-1050/18/3/1486>
- [15] Giudici, Paolo, and Lunshuai Wu. "Sustainable artificial intelligence in finance: Impact of ESG factors." *Frontiers in Artificial Intelligence* 8 (2025): 1566197. Available: <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1566197/full>
- [16] Musleh Al-Sartawi, Abdalmutaleb MA, Khaled Hussainey, and Anjum Razzaque. "The role of artificial intelligence in sustainable finance." *Journal of Sustainable Finance & Investment* (2022): 1-6. Available: <https://www.tandfonline.com/doi/full/10.1080/20430795.2022.2057405>
- [17] Dipierro, Anna Rita, Fernando Jimenéz Barrionuevo, and Pierluigi Toma. "Predicting ESG controversies in banks using machine learning techniques." *Corporate Social Responsibility and Environmental Management* 32, no. 3 (2025): 3525-3544. Available: <https://onlinelibrary.wiley.com/doi/full/10.1002/csr.3146>
- [18] Lim, Tristan. "Environmental, social, and governance (ESG) and artificial intelligence in finance: State-of-the-art and research takeaways." *Artificial Intelligence Review* 57, no. 4 (2024): 76. Available: <https://link.springer.com/article/10.1007/s10462-024-10708-3>
- [19] Dewasiri, Narayanage Jayantha, Dunusinghe G. Dharmarathna, and Mrinalini Choudhary. "Leveraging artificial intelligence for enhanced risk management in banking: A systematic literature review." *Artificial intelligence enabled management: An emerging economy perspective* (2024): 197-213. Available: <https://books.google.co.in/books?hl=en&lr=&id=wgkFEQAAQBAJ&oi=fnd&pg=PA197>
- [20] Pluskota, Przemysław, Kamila Słupińska, Agata Wawrzyniak, and Barbara Wąsikowska. "The Application of Artificial Intelligence (AI) in the Implementation of ESG-Oriented Sustainable Development Strategies in the Banking Sector: A Case Study." *Sustainability* 18, no. 2 (2026): 732. Available: <https://www.mdpi.com/2071-1050/18/2/732>