رفاد REFAAD
www.refaad.com

Global Journal of Economics and Business (GJEB)
المجلة العالمية للاقتصاد والأعمال
Journal Homepage: https://www.refaad.com/Journal/Index/2
E-ISSN 2519-9293 | P-ISSN 2519-9285
Vol. 16, Issue 3, 2026, pp. 373 - 385

Measuring ESG Disclosure Quality Using Bilingual Natural Language Processing: A Proposed Methodological Framework for Large-Cap Saudi Tadawul-Listed Companies

Saleh Mohammed Baqader[1*]
[1] Associate professor of Accounting, College of Business and Economics Umm Al-Qura University, The Kingdom of Saudi Arabia.
[*] Corresponding Author: Saleh Baqader (smbaqader@uqu.edu.sa)

قياس جودة إفصاحات الحوكمة البيئية والاجتماعية والمؤسسية (ESG) باستخدام معالجة اللغة الطبيعية الثنائية اللغة: إطار منهجي مقترح للشركات السعودية الكبيرة المُدرَجة في السوق المالية (تداول)

صالح محمد باقادر[1*]
[1] أستاذ مشارك في المحاسبة- كلية الأعمال والاقتصاد- جامعة أم القرى- المملكة العربية السعودية.
*الباحث المراسل: صالح باقادر (smbaqader@uqu.edu.sa)

Received استلام البحث	Revised مراجعة البحث	Accepted قبول البحث
2026/5/9	2026/5/31	2026/6/23

DOI: https://doi.org/10.31559/GJEB2026.16.3.9

Abstract:
Firms' disclosure of environmental, social, and governance (ESG) information gives invaluable insights to investors and helps stock markets function effectively. Existing research offers useful techniques for measuring the quantity of ESG disclosure, but there has been too little focus on quality indicators. Unresolved problems also exist in dealing with bilingual source material. To address these issues, this paper presents a methodological framework for measuring the quality of ESG disclosures in a bilingual (English–Arabic) setting. It designs a natural language processing (NLP) pipeline capable of analyzing ESG disclosure quality in the annual reports of companies listed on the Saudi stock exchange, the Tadawul. The ESG_C framework processes Arabic- and English-language content separately, since both are used in Saudi corporate reports. It measures disclosure quality using a weighted index comprising verifiability, sector specificity, quantitative clarity, and standards alignment. The framework is used to compile a structured corpus of 250 annual reports from 50 Tadawul-listed companies (21.6% of listed firms) spanning the fiscal years 2019–2023. The sample includes only firms with complete, machine-readable reports and a market value of at least SAR 500 million, leading to a large-cap focus and the exclusion of 8 of the 21 industry groups represented on the Tadawul. The study does not empirically test the entire NLP pipeline, validate the combined index, or present empirical results on ESG disclosure quality, but it reports an initial sentence-level ESG relevance classifier with an F1 score of 0.87 on a 300-sentence hold-out set, offering a transparent measurement tool and a well-documented corpus design for future validation.

Keywords: *ESG disclosure quality; natural language processing; Saudi exchange; Tadawul; sustainability reporting; Arabic NLP; AraBERT; methodological framework; annual report corpus; bilingual report processing.*

الملخص:
يوفّر إفصاح الشركات عن معلومات الجوانب البيئية والاجتماعية والمؤسسية (ESG) معلومات أساسية للمستثمرين، تُرشِد قراراتهم وتساعد أسواق الأسهم على العمل بكفاءة. غير أن البحوث المتعلقة بإفصاحات ESG كثيراً ما قاست الكمّ بدلًا من الجودة، مما يحدّ من قيمتها. وعلاوةً على ذلك، لا تقدّم الأساليب القائمة سوى حلول محدودة لتحليل المواد المصدرية الثنائية اللغة. ولذلك تقترح هذه الورقة إطارًا منهجيًا لقياس جودة إفصاحات ESG في سياق ثنائي اللغة (إنجليزي-عربي). وتصمّم الورقة مسار معالجة للغة الطبيعية (NLP) قادرًا على تحليل جودة إفصاحات ESG في التقارير السنوية للشركات المُدرَجة في السوق المالية السعودية (تداول). ويستخدم إطار ESG_C مكوّنات معالجة منفصلة للمحتوى باللغتين العربية والإنجليزية، نظرًا لاستخدامهما معًا في تقارير الشركات السعودية. ويقيس الإطار جودة الإفصاح باستخدام مؤشر مرجَّح رباعي الأبعاد يتألف من قابلية التحقق، والتخصّص القطاعي، والوضوح الكمّي، والمواءمة مع المعايير. ويُستخدَم الإطار الثنائي اللغة لتجميع مدوّنة مُنظَّمة من 250 تقريرًا سنويًا لخمسين شركة كبيرة مُدرَجة في تداول (21.6% من الشركات المُدرَجة) تغطي السنوات المالية 2019–2023. ولا تشمل العينة سوى الشركات التي لديها تقارير كاملة قابلة للقراءة آليًا وقيمة سوقية لا تقل عن 500 مليون ريال سعودي، مما أفضى إلى التركيز على الشركات الكبيرة واستبعاد 8 من أصل 21 مجموعة صناعية ممثَّلة في تداول. ولا تختبر الدراسة تجريبيًا مسار معالجة اللغة الطبيعية بأكمله، ولا تتحقق من المؤشر المُجمَّع، ولا تعرض نتائج تجريبية عن جودة إفصاحات ESG، لكنها تورد مُصنِّفًا أوليًا لمدى صلة الجمل بـESG على مستوى الجملة بدرجة F1 تبلغ 0.87 على مجموعة اختبار محتجَزة قوامها 300 جملة، بما يوفّر أداة قياس شفافة وتصميم مدوّنة موثّقاً جيداً للتحقق منه مستقبلًا.

الكلمات المفتاحية: *معالجة اللغة الطبيعية؛ السوق المالية السعودية؛ تداول؛ تقارير الاستدامة؛ معالجة اللغة الطبيعية العربية؛ AraBERT؛ إطار منهجي؛ مدوّنة التقارير السنوية؛ معالجة التقارير الثنائية اللغة.*

Citation
Baqader, S. M. (2026). Measuring ESG Disclosure Quality Using Bilingual Natural Language Processing: A Proposed Methodological Framework for Large-Cap Saudi Tadawul-Listed Companies. *Global Journal of Economics and Business*, 16(3), 373-385. https://doi.org/10.31559/GJEB2026.16.3.9

1 Introduction

Firms' disclosures of environmental, social, and governance (ESG) information help drive investors' decisions across a wide variety of regulatory and cultural contexts, offering key guidance on responsible investments (United Nations Principles for Responsible Investment, 2022). As investors increasingly rely on credible ESG information, academic attention is shifting from measuring the mere existence of disclosures to evaluating their quality. This work is ongoing, since existing methods have not succeeded in fully measuring quality (Berg et al., 2022; Hummel & Schlick, 2016) and important tools such as ESG ratings and manual disclosure scores can have unclear underlying methods. They can also be difficult to replicate, especially in bilingual reporting environments. In this context, a validated NLP-based approach can offer valuable methodological improvements. Moreover, while these methodological challenges persist, the shift towards frameworks focusing on quality has been given added impetus by evidence of a divergence between firms' ESG disclosure volume and underlying performance (Cho et al., 2012).

However, some markets present particular challenges for measuring the quality of ESG disclosures. For example, there are unique difficulties in applying accepted frameworks in the Gulf Cooperation Council (GCC) member states, where regulations and practices are evolving and corporate documents use both Arabic and English, sometimes in the same reports. As the largest equity market in the region, the Tadawul in Saudi Arabia provides an ideal context for studying frameworks that are better fitted to a bilingual, evolving, complex ESG disclosure context. The country's Capital Market Authority (CMA) has already introduced tighter regulations and updated its ESG reporting guidance for main market issuers (Capital Market Authority Saudi Arabia, 2021) as part of the government's Vision 2030 programme, whose aims including enhancing transparency and aligning regulations with global standards.

The additional ESG information resulting from these changes provides valuable potential data for investors and researchers but highlights the lack of suitable tools that can handle the bilingual material. A new, transparent method is therefore needed to simplify the process of assessing the quality of ESG disclosures, making it easier to use the information effectively for research as well as investment purposes. This study proposes a bilingual natural language processing (NLP) pipeline capable of harvesting and analyzing ESG disclosure information, creating a practical tool that is currently lacking. In turn, this knowledge would better inform investors' decisions and would be especially valuable in the country's evolving regulatory context. The study does not include a formal systematic review of the literature. However, the consulted literature indicates that a bilingual NLP-based tool for evaluating ESG disclosure quality in Saudi annual reports across various sectors and years is still underdeveloped.

This study therefore aims to fill this methodological gap in the literature. It has three research objectives. Firstly, it seeks to design a bilingual NLP pipeline to measure ESG disclosure quality in the annual reports of firms listed on the Tadawul. Secondly, it aims to construct a documented corpus of 250 annual reports with full company-level, sector-level, and source documentation. Finally, it aims to characterize the scope, coverage, and selection bias of the sample relative to all the firms listed on Tadawul.

The research sample comprises 50 of the 232 firms listed on the Tadawul (21.6% of the total number; approximately 68% by market capitalization in 2023). These 50 firms represent 13 of the 21 industry groups present on the Tadawul. The sample is limited to large-cap issuers by an inclusion criterion of a market capitalization threshold of SAR 500 million market. The threshold was chosen in the study design, not based on empirical calibration, and thus limits the analysis to larger companies with better disclosure and reporting capacity.

2 Literature Review

Previous literature on corporate sustainability reporting concentrated on volume-based indicators such as page counts, disclosure indices, and binary presence measures (Deegan, 2002; Gray et al., 1995). More recent studies have sought to expand this purely statistical approach into quality indicators, but not without obstacles. For example, the use of third-party ESG ratings was criticized on several grounds, including the risk that different raters might apply different standards (Berg et al., 2022), weakening the reliability of the indicators. Other authors (Clarkson et al., 2008) have categorized disclosures into hard and soft types, the former comprising data that is quantitative and/or verifiable, and the latter referring to qualitative and/or generic information. This distinction between hard and soft disclosures established a theoretical foundation for measuring quality. Certain other characteristics of ESG disclosures are also considered to offer clear

markers of quality. For example, sector specificity and verifiability were identified by Hummel and Schlick (2016) as the most helpful dimensions for assessing quality.

Prior studies have also proposed and assessed a variety of methods for analyzing financial reports using lexicon-based methods. Loughran and McDonald (2016) noted that relying on textual analysis rather than quantitative methods was less precise, more nuanced, and prone to pitfalls. Later work by Bingler et al. (2022) demonstrated that fine-tuned BERT models classified climate-related disclosures at sentence level with substantially higher precision than dictionary methods. For Arabic text, the AraBERT model (Antoun et al., 2020) and CAMeL Tools (Obeid et al., 2020) provide the necessary infrastructure for modern standard Arabic NLP. However, to the best of the authors' knowledge, no prior studies have applied these tools to analyzing ESG disclosure in annual reports in Arabic.

Given the morphological richness of Arabic ESG terminology, the use of transformer models in this context should be considered exploratory until external validation and inter-rater reliability testing are completed. It should also be noted that commercial ESG scoring systems and products designed for use in multilingual contexts are available. These may function in broadly similar ways to the proposed framework. However, as commercial products, they cannot be reproduced in other settings. The framework proposed in this study therefore offers a transparent academic alternative.

Some prior research has highlighted the importance of measuring the quality of the information disclosed by firms. For example, Michelon et al.'s (2015) study on corporate social responsibility (CSR) concluded that firms' disclosures sometimes offered symbolic compliance rather than credible content, justifying skepticism towards such information. Other researchers have studied different aspects of disclosures by Saudi firms and other member states of the GCC. For example, Alotaibi and Hussainey (2016) examined the determinants of CSR disclosure quantity and quality for non-financial Saudi-listed firms, while Bamahros et al. (2022) studied corporate governance mechanisms and ESG reporting among Saudi-listed companies. A positive association between corporate governance quality and CSR disclosure has also been demonstrated (Habbash, 2016). Other researchers found that initiating voluntary CSR reporting reduced the cost of equity capital (Dhaliwal et al., 2011). In the broader GCC context, ESG performance has been shown to influence corporate financial performance (Abdullah et al., 2025). Haniffa and Cooke (2005) established that cultural governance dimensions shaped reporting behavior in Malaysia while Ioannou and Serafeim (2019) have provided international evidence that evolving reporting expectations tend to improve disclosure behavior. Manual disclosure scores and commercial ESG products remain central comparative tools for evaluating the usefulness and validity of any future NLP-based disclosure measure. No prior study could be identified that has deployed an NLP-based, bilingual ESG disclosure quality measurement instrument for Saudi annual reports in a longitudinal multi-sector sample.

3 Methodology

3.1 Sample Construction

To begin the process of developing an NLP pipeline suitable for assessing ESG disclosure quality on the Saudi market, criteria were identified that would ensure a consistent five-year panel within a selected large-cap segment and with comparable document inputs. This led to a three-stage screening procedure that was applied to the firms listed on the Tadawul in January 2024 (N = 232) to ensure the data gathered was consistent and suitable for the purposes of the study (see Table 1). The first criterion was for the firms to have been continuously listed from January 2019 to December 2023, so the study was based on disclosures over a consistent period for all firms. This resulted in the exclusion of 63 firms that did not meet this criterion. Next, it was required that all of the firms' annual reports for the same period should be available and machine-readable, ensuring their suitability for the NLP tool. This meant that firms with missing reports, inaccessible files, or non-processable PDF formats were excluded. A further 82 firms were excluded for these reasons. Finally, a minimum market capitalization of SAR 500 million as of December 2023 was set. This threshold ensured that the study only included companies with adequate investor relations (IR) infrastructure. Another 37 firms were excluded. After applying the screening process, the final sample thus comprised 50 of the initial target population of 232 firms. The sample encompassed 13 of the industry groups represented on the Tadawul. It enabled a total of 250 firm-year observations to be gathered.

Table 1: Screening Process for Tadawul-listed Firms

Stage	Criterion	Excluded	Remaining
Initial Universe	All Tadawul listings, January 2024	—	232
Screen 1	Continuous listing 2019–2023	63	169
Screen 2	All five annual reports machine-readable	82	87
Screen 3	Market capitalization ≥ SAR 500M (December 2023)	37	50
Final Sample	13 Tadawul industry groups; 250 firm-years	—	**50**

3.2 Annual Report Collection

Annual reports were collected from the 50 firms in the sample for the five years of the study period, 2019-2023. Most were available on the Saudi Exchange Disclosure Portal (Saudi Exchange [Tadawul], 2026). This constituted a primary source of annual reports, enabling the identification of 63 URLs that were verified by direct access. Another 40 verified reports were downloaded from corporate websites dealing with IR affairs, constituting a secondary source. Finally, accessing the Argaam Plus financial archive (Argaam, 2026) resulted in 18 firm-year records. This tertiary source completed the process, leading to a final corpus of 250 annual reports. Although the reports were verified through available sources, using portal, IR, and archive copies means that version differences cannot be completely ruled out. Future studies should improve reproducibility by recording retrieval dates and storing archive checksums.

The vast majority (88%) of the corpus documents were natively digital PDFs, with the remaining 12% requiring the use of optical character recognition (OCR) tools. The length of the documents varied widely, with the shortest containing fewer than 16,000 words, while the longest had nearly 60,000 words. The fact that the corpus contained documents in both English and Arabic reinforced the need for the bilingual pipeline. For example, larger firms trying to attract international investors (e.g., Saudi Aramco, SABIC, STC, SNB) had published comprehensive English-language sustainability reports alongside their Arabic-language statutory annual reports, while the medium-sized and smaller firms tended to use only Arabic. The manuscript does not include a detailed year-by-year breakdown of report language categories, including Arabic-only, English-only, and mixed-language reports, which could be addressed in future work. Moreover, it should be noted that potential OCR errors also represent a significant limitation, because misread numerals or negation terms could affect ESG classification and scores. Full details of the corpus characteristics are shown in Table 2, while Table 3 shows the distribution between sectors.

Table 2: Characteristics of Corpus of Annual Reports

Characteristic	Value
Total number of reports	250
Natively digital PDFs	220 (88%)
Image-based PDFs (OCR required)	30 (12%)
Mean document length (words)	37,804
Standard deviation	14,200
Minimum document length (words)	15,714
Maximum document length (words)	59,652

Note: OCR = optical character recognition.

Table 3: Sample Composition by Sector

Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG
Measuring ESG	Measuring ESG	Measuring ESG

3.3 ESG Disclosure Quality Framework

The ESG disclosure quality framework operationalized ESG disclosure quality through four dimensions: verifiability, sector specificity, quantitative clarity, and standards alignment (see Table 4). Verifiability refers to the proportion of ESG-relevant sentences within the report that contained quantitative data or references to

external assurances. Next, sector specificity is the cosine similarity between the firm's ESG disclosure vector and the material topics for the firm's industry defined by the Sustainability Accounting Standards Board (SASB). Since the SASB standards were developed in a different institutional setting, the sector specificity component should be interpreted as an approximation of international materiality alignment rather than as a confirmed Saudi market-specific materiality map. Quantitative clarity captures the density of language used in reports that sets goals for the firm, such as numerical targets, baselines, key performance indicators (KPIs), and year-on-year comparisons. To prevent double counting, verifiability and quantitative clarity were kept conceptually distinct: verifiability involved evidence or assurance, while quantitative clarity involved target, baseline, KPI, and comparison structures. Finally, the dimension of standards alignment refers to the frequency of explicit references contained in the reports to Global Reporting Initiative (GRI) Standards, recommendations from the Task Force on Climate-related Financial Disclosures (TCFD), SASB metrics, or the United Nations Sustainable Development Goals (SDGs). The four dimensions were weighted differently in the framework, as also shown in Table 4. Verifiability was weighted at 25%, sector specificity and quantitative clarity were both weighted at 30%, and finally standards alignment was weighted at 15%. The following formula was used to reflect this weighting:

$$ESG\_C = 0.25 \times ESG\_V + 0.30 \times ESG\_Sp + 0.30 \times ESG\_Q + 0.15 \times ESG\_M \quad (1)$$

The selection of these weights was based on the theoretical relevance of quantitative specificity and sector-relevance in the ESG quality literature (Clarkson et al., 2008; Hummel & Schlick, 2016). However, it is acknowledged that they have not been derived from expert surveys or empirical calibration and should be subjected to sensitivity analysis in future validation studies. To assess the weighting sensitivity, future validation will be required to compare equal-weighted, expert-weighted, and data-driven specifications and determine whether firm-year rankings are significantly impacted. The study drew on several theories to justify the four dimensions incorporated in the framework. Signaling theory (Spence, 1973) underpins the importance of featuring the dimension of verifiability. It suggests that credible disclosures that can be confirmed externally will be a better guide to substantive ESG performance better than the use of symbolic narrative reporting. The research also draws on stakeholder theory (Freeman, 1984), which recognizes the likely difference between the different effects of ESG disclosure in the varying sectors represented in the sample. This variation based on sector specificity occurs as stakeholder expectations are likely to differ significantly depending on the industry. There are also sector-specific sustainability risks that investors are likely to take into account. As a result, stakeholder theory suggests that the findings are unlikely to be homogenous across sectors and explains the reasons based on expectations and risks.

Additional theoretical support was drawn from the hard disclosure framework (Clarkson et al., 2008). This supported quantitative clarity by prioritizing measurable, target-oriented, and decision-useful disclosures over vague qualitative statements. Finally, legitimacy theory (Suchman, 1995) supported the dimension of standards alignment, offering understanding why it is so important for firms to meet international norms. The theory suggests that firms will wish to seen as meeting internationally recognized disclosure frameworks to emphasize their own legitimacy as institutions, and to demonstrate that they are acting in conformity with regulatory and societal expectations. However, legitimacy theory also suggests that firms may cite standards in order to signal compliance rather than actually demonstrate it. Therefore, references to standards should be viewed with caution unless supported by verifiable and quantitative disclosures.

Table 4: Theoretical and Methodological Structure of ESG_C: Component Definitions, Theoretical Foundations, and Weighting Scheme

Dimension	Code	Definition	Theoretical Basis	Weight
Verifiability	ESG_V	Proportion of ESG-relevant sentences containing quantitative data or external assurance references	Signaling (Spence, 1973)	25%
Sector specificity	ESG_Sp	Cosine similarity between firm ESG disclosure vector and SASB-defined material topics for the firm's industry	Stakeholder theory (Freeman, 1984)	30%
Quantitative clarity	ESG_Q	Density of goal-setting language: numerical targets, baselines, KPIs, and year-on-year comparisons	Hard disclosure (Clarkson et al., 2008)	30%
Standards alignment	ESG_M	Frequency of explicit references to GRI standards, TCFD recommendations, SASB metrics, or UN SDGs	Legitimacy theory (Suchman, 1995)	15%

3.4 Natural Language Processing Pipeline Design

After identifying the sample and building the corpus, the next stage was designing the NLP pipeline. First, the text was extracted from the documents, using two different methods depending on whether the PDFs were natively digital or image-based. For the former group, the processing was conducted using PyMuPDF (v1.23), with the original layout preserved. For the latter group, Tesseract OCR (v5.3, Arabic tessdata) was used. To ensure adequate accuracy had been achieved for the OCR processing, a manual validation check was conducted using a 500-page subset of the corpus. This estimated the character accuracy at 94.3%, implying an error rate of approximately 5.7%, which was considered acceptable. The pipeline treats tables, figures, and footnotes primarily as text-extraction challenges rather than as separate structured elements; therefore, ESG information contained only in non-textual components may not have been captured. This text extraction process was followed by text segmentation and normalization in Stage 2 of the pipeline design. Stanza (v1.7, Arabic model) and spaCy (v3.6, English model) were used to detect the sentence boundaries. The Arabic text was normalized using the Farasa toolkit and CAMeL Tools for structural disambiguation.

For Stage 3 of the design process, the sentences contained in the corpus were classified based on their relevance to the topic of ESG disclosure. To enable this classification to take place, a bilingual dictionary of ESG terms was created. These terms were drawn from the GRI Standards taxonomy (Global Reporting Initiative, 2021), SASB sector-specific metrics (Sustainability Accounting Standards Board, 2018), and TCFD recommended disclosures (Task Force on Climate-related Financial Disclosures, 2021). The final dictionary contained 300 terms, of which 162 were Arabic and 138 were English. These terms were then placed in sub-lexicons depending on whether they related to environmental, social, or governance content. The Arabic translations were validated by a bilingual expert in the field. Contextual disambiguation was also required to ensure all the terms were interpreted as intended in their original context in the documents. This was carried out using an AraBERT model (Antoun et al., 2020), fine-tuned on 1,200 manually labeled sentences using CAMeL Tools (Obeid et al., 2020). To assess the performance of the sentence-level ESG relevance classifier, the study used a validation subset of 300 manually labeled sentences. For this subset, the AraBERT classifier achieved an F1 score of 0.87 for binary ESG relevance classification. It should be noted that this validation covered the sentence-level classification performance but was not a comprehensive validation of the ESG_C framework or its sub-index weighting structure. The current validation evidence should be considered as preliminary since the design lacks cross-validation, held-out firm or year testing, inter-rater reliability statistics, or error patterns by ESG subcategory.

In Stage 4, the study calculated scores for each sub-index. For example, the score for verifiability (coded as ESG_V) was calculated as the proportion of ESG-classified sentences containing at least one numerical token. For sector specificity (ESG_Sp), the score reflected the cosine similarity to the SASB sector keyword set. Quantitative clarity (ESG_Q) was scored on the basis of the density of goal-related phrases per 1,000 ESG words. Finally, for standards alignment (ESG_M), the score reflected the number of explicit standard references normalized by total ESG sentences. The final stage involved composite aggregation. All the sub-indices were min-max standardized to [0,1] within the sample. ESG_C was computed as the weighted sum per firm per year, yielding 250 firm-year records.

3.5 Proposed Empirical Model and Illustrative Example

For future empirical application, once the NLP pipeline validation is complete, the study proposes the following exploratory two-way fixed effects panel model, subject to later robustness testing:

$$\mathbf{ESG\_C_{it} = \alpha + \beta_1 \ln(Size_{it}) + \beta_2 ROA_{it} + \beta_3 DocLen_{it} + \gamma_{sector} + \delta_{year} + \varepsilon_{it}} \quad \mathbf{(2)}$$

No coefficients, standard errors, p-values, or inferential statistics are estimated or reported in this paper as the framework has not yet been applied empirically. However, the framework's intended operation can be demonstrated using a hypothetical passage from an annual report. This passage is illustrative only and should not be interpreted as evidence drawn from the corpus. For instance, if a document cited a quantitative environmental target aligned with a recognized standard, the pipeline would assign this a high verifiability (ESG_V) score, as the sentence contained numerical data that is verified. Similarly, if the terminology closely matched the firm's sector-specific SASB keyword set, the passage would achieve a high score for sector specificity (ESG_Sp). For content mentioning a baseline, current value, and forward target but no specific GRI/TCFD indicator, the outcome would be a high ESG_Q score but only a partial ESG_M score, as the external indicator was not explicitly cited. Overall, detailed content of this type would achieve a high

composite score. Conversely, a passage containing vague commitments but lacking statistics, sector-specific terms, targets, or references to standards would receive very low scores across all four dimensions, resulting in a low composite score. While based on hypothetical content, this contrast in scores for the two types of content demonstrates how the framework is designed to differentiate between substantive ESG disclosures and generic or imprecise information. It is hoped that future empirical research will evaluate the model using firm-clustered standard errors, heteroskedasticity-robust inference, alternative fixed effect structures, and broader control sets, such as firm age, board independence, ownership structure, and regulatory exposure.

4 Transparency Report for the Study Sample

4.1 Full Sample Disclosure

Although this study does not empirically implement the framework, it provides full disclosure of the sample composition, market coverage, and selection biases to ensure transparency. Table 5 presents details on 50 firms used in the study, including their names, sectors, industries, the number of reports (five in all cases), and report sources.

Table 5: Complete Sample of 50 Tadawul-Listed Companies (Annual Reports 2019–2023)

No.	Ticker	Company Name	GICS Sector	Tadawul Industry	Reports	Source
1	1120	Al Rajhi Bank	Banks	Banks	5	Company IR Website
2	1150	Alinma Bank	Banks	Banks	5	Saudi Exchange Portal
3	1080	Arab National Bank	Banks	Banks	5	Company IR Website
4	1140	Bank Al Bilad	Banks	Banks	5	Argaam Plus Archive
5	1020	Bank Al Jazira	Banks	Banks	5	Argaam Plus Archive
6	1050	Banque Saudi Fransi	Banks	Banks	5	Company IR Website
7	1010	Riyad Bank	Banks	Banks	5	Company IR Website
8	1060	Saudi British Bank (SABB)	Banks	Banks	5	Company IR Website
9	1030	Saudi Investment Bank (SAIB)	Banks	Banks	5	Saudi Exchange Portal
10	1180	Saudi National Bank (SNB)	Banks	Banks	5	Company IR Website
11	8010	Tawuniya (NCCI)	Insurance	Insurance	5	Company IR Website
12	8120	Bupa Arabia	Insurance	Insurance	5	Company IR Website
13	8060	Walaa Cooperative Insurance	Insurance	Insurance	5	Saudi Exchange Portal
14	8050	Gulf Union Al Ahlia Insurance	Insurance	Insurance	5	Argaam Plus Archive
15	2222	Saudi Aramco	Energy	Energy	5	Company IR Website
16	2380	Petro Rabigh	Energy	Energy	5	Company IR Website
17	2010	SABIC	Materials	Materials	5	Company IR Website
18	4200	Aldrees Petroleum	Energy	Energy	5	Saudi Exchange Portal
19	1211	Maaden (Saudi Arabian Mining)	Materials	Materials	5	Company IR Website
20	2310	SIPCHEM	Materials	Materials	5	Argaam Plus Archive
21	2060	TASNEE	Materials	Materials	5	Argaam Plus Archive
22	2330	Advanced Petrochemical	Materials	Materials	5	Saudi Exchange Portal

23	2240	Zamil Industrial	Materials	Capital Goods	5	Argaam Plus Archive
24	2350	Saudi Kayan Petrochemical	Materials	Materials	5	Argaam Plus Archive
25	2290	YANSAB	Materials	Materials	5	Argaam Plus Archive
26	2020	SABIC Agri-Nutrients	Materials	Materials	5	Company IR Website
27	1603	Saudi Cable Company	Materials	Capital Goods	5	Saudi Exchange Portal
28	3030	Saudi Cement Company	Materials	Materials	5	Argaam Plus Archive
29	3040	Qassim Cement (QACCO)	Materials	Materials	5	Saudi Exchange Portal
30	2270	Sadafco (Saudia Dairy)	Consumer Staples	Food, Beverages & Tobacco	5	Argaam Plus Archive
31	5110	Saudi Electricity Company (SEC)	Utilities	Utilities & Energy Services	5	Company IR Website
32	2081	Marafiq Power & Water	Utilities	Utilities & Energy Services	5	Argaam Plus Archive
33	2083	TAQA (Saudi Arabian Water)	Utilities	Utilities & Energy Services	5	Argaam Plus Archive
34	2082	ACWA Power	Utilities	Utilities & Energy Services	5	Company IR Website
35	4002	Mouwasat Medical Services	Health Care	Health Care Equip. & Svcs	5	Argaam Plus Archive
36	4013	Dr. Sulaiman Al-Habib (HMG)	Health Care	Health Care Equip. & Svcs	5	Company IR Website
37	4004	Dallah Health	Health Care	Health Care Equip. & Svcs	5	Saudi Exchange Portal
38	1214	Shaker Group	Industrials	Capital Goods	5	Argaam Plus Archive
39	4030	Bahri (National Shipping Co.)	Industrials	Transportation	5	Company IR Website
40	4031	Saudi German Hospitals	Industrials	Capital Goods	5	Saudi Exchange Portal
41	2280	Almarai Company	Consumer Staples	Food, Beverages & Tobacco	5	Company IR Website
42	2050	Savola Group	Consumer Staples	Food, Beverages & Tobacco	5	Company IR Website
43	2300	NADEC	Consumer Staples	Food, Beverages & Tobacco	5	Argaam Plus Archive
44	4161	BinDawood Holding	Consumer Staples	Retailing	5	Company IR Website
45	4190	Jarir Marketing Group	Consumer Disc.	Retailing	5	Company IR Website
46	4210	Saudi Research & Media (SRMG)	Consumer Disc.	Media & Entertainment	5	Company IR Website
47	7203	ELM Company	Technology	Technology	5	Saudi Exchange Portal
48	7202	Solutions by STC	Technology	Technology	5	Saudi Exchange Portal
49	7010	STC (Saudi Telecom Company)	Telecom.	Telecom Services	5	Company IR Website
50	7020	Mobily (Etihad Etisalat)	Telecom.	Telecom Services	5	Company IR Website

Notes: Source platforms – Saudi Exchange Disclosure Portal Saudi Exchange [Tadawul], 2026): 63 verified URLs; Corporate IR websites: 40 verified direct downloads; Argaam Plus archive: 18 records. GICS = Global Industry Classification Standard.

4.2 Sector Coverage

As noted in Section 3, the sample of 50 firms represented 13 of the 21 industry groups present on the Tadawul. To ensure full transparency, Table 6 sets out which of the industry groups was present in or absent from the sample, and the number of firms from each industry group. It also provides notes on the level of sector coverage represented by the number of firms in the sample. The study acknowledges here that the

absence of eight industry groups highlights that the findings of this study cannot be generalized to the whole of the Tadawul. In addition, insurance-sector coverage is limited rather than comprehensive, and the lack of coverage across several sectors prevents full cross-sector generalization.

Table 6: Tadawul Sector Coverage Audit

Status	Tadawul Industry Group	Firms	Coverage Note
Covered	Banks	10	10 of ~12 listed banks
Covered	Insurance	4	4 of ~34 listed insurance firms — undersampled
Covered	Energy	3	3 of ~5 listed energy companies
Covered	Materials	9	Chemicals, cement, mining, cable
Covered	Capital Goods	4	Zamil Industrial, Saudi Cable, Shaker Group, Saudi German Hospitals
Covered	Transportation	1	Bahri only — others excluded by market cap
Covered	Food, Beverages & Tobacco	4	Almarai, Savola, NADEC, Sadafco
Covered	Retailing	2	BinDawood, Jarir
Covered	Media & Entertainment	1	SRMG only
Covered	Health Care Equip. & Services	3	Mouwasat, HMG, Dallah
Covered	Utilities & Energy Services	4	SEC, Marafiq, TAQA, ACWA Power
Covered	Technology	2	ELM, Solutions by STC
Covered	Telecom Services	2	STC, Mobily
Absent	Real Estate Mgmt & Development	0	~30 REITs absent — no reports collected
Absent	Diversified Financials	0	Investment/leasing firms — excluded by market cap filter
Absent	Pharma, Biotech & Life Sciences	0	~8 firms — incomplete report availability
Absent	Commercial & Professional Services	0	~12 firms — no verified report URLs
Absent	Consumer Services	0	Hotels, restaurants (~10 firms) — all excluded
Absent	Household & Personal Products	0	~6 firms — excluded by market cap filter
Absent	Food & Staples Retailing	0	~4 firms — excluded by market cap filter
Absent	Automobiles & Components	0	~3 firms — excluded by market cap filter

4.3 Market Coverage

This study involved only 50 of the 232 firms that were listed on the Tadawul as of January 2024 (see Section 3.1). This further highlights that the full Saudi stock market is not covered and the findings cannot be generalized to the whole index. A high market cap does not guarantee representativeness, as firm size and reporting capacity can systematically affect disclosure quality. Table 7 sets out the number of firms included and excluded and market coverage by number and market capitalization.

Table 7: Market Coverage Summary

Metric	Value
Total listed companies, Saudi Exchange (January 2024)	232
Companies in this study	50
Market coverage by firm count	21.6%
Market coverage by market capitalization (2023 estimate)	~68%
Companies excluded (total)	182

4.4 Evidence-Based Exclusion Reasons

As detailed in the Methodology (see Section 3.1), the study employed a three-stage screening process to ensure a balanced panel across the study period. This resulted in the exclusion of 182 firms and eight industry sectors (see Tables 5 and 6). As detailed in Section 3.1, 63 of the 232 firms listed in January 2024 were excluded because they had been listed after January 2019, i.e. after the start of the study period. This would have resulted in an incomplete five-year panel. Secondly, another 82 firms were excluded from the sample because they had missing, inaccessible, or non-machine-readable annual reports. This included issues such as broken IR website links, PDFs that had been scanned in Arabic without a searchable text layer, documents that were inaccessible because they were password-protected, and the absence of any CMA-registered IR presence. Finally, 37 additional firms were excluded from the sample as they did not meet the market capitalization threshold. The threshold was set at SAR 500 million to ensure that only companies with sufficient IR

infrastructure were included. The manuscript lacks a systematic failure log for each firm and year. Hence, future research should develop such a record to support replication.

4.5 Bias Disclosure

The selection criteria adopted for this study resulted in several areas of concern relating to imbalances within the resultant panel. This included a bias towards large firms, particularly as a result of the market capitalization threshold that was set. In addition, some industry sectors (namely, banks and materials) were over-represented in the sample, while insurance was severely under-represented and eight other sectors were absent, including the entire real estate investment trust (REIT) and real estate group. The study does not evaluate the extent of these biases using weighting methods or counterfactual inclusion tests. At the same time, the 2019–2023 period includes the COVID-19 pandemic and oil-price shocks, which may have led to temporal effects in corporate reporting behavior. Table 8 presents the results of the assessment of the bias within the sample.

Table 8: Bias Assessment of Study Sample

Bias Dimension	Assessment	Evidence
Large-firm bias	Yes	Explicit SAR 500M market cap filter. 10 largest firms account for approximately 60% of total corpus word volume.
Sector over-representation	Banks 20%; Materials 22%	Banks represented 10 of the 50 firms in the sample (20%) but only represent approximately 5% of Tadawul listings.
Sectors entirely absent	8 Tadawul groups	The missing industry groups were as follows: Real Estate/REITs, Diversified Financials, Pharma/Biotech, Commercial & Professional Services, Consumer Services, Household Products, Food Staples Retailing, Automobiles.
Insurance under-representation	Severe (11.8%)	4 of approximately 34 listed insurance companies (11.8%), all large-cap.

4.6 Scope Statement

In summary, the scope of this study is restricted to 50 large-cap Tadawul-listed companies with machine-readable annual reports available for the fiscal years 2019–2023. This represents 13 of the 21 industry groups represented on the Tadawul. Small- and mid-cap issuers were excluded by the SAR 500 million market capitalization threshold, and eight industry groups are entirely absent from the sample. The study's results cannot therefore be generalized to the entire Saudi market. They are also not generalizable to other GCC states or other emerging market contexts. Future researchers could conduct similar studies in these contexts to validate the framework's suitability for use.

5 Discussion

The bilingual framework proposed and developed by this research marks an important step forward in measuring the quality of ESG disclosures in complex, bilingual, or emerging economy contexts. While developed specifically in the Saudi context, it has strong potential for use in numerous other settings where traditional frameworks for measuring the ESG disclosure quality may not provide a good fit. By operationalizing four dimensions of quality (verifiability, sector specificity, quantitative clarity, and standards alignment), and building on the key distinction between hard and soft disclosures noted by prior researchers (Clarkson et al., 2008), the framework also supports progress from quantity-based approaches to quality-based analysis. In this regard, it was directly informed by the prior literature suggesting verifiability and sector specificity were key dimensions of assessing quality (Hummel & Schlick, 2016).

This disaggregation of the dimensions of ESG disclosure quality enables a far more meaningful and nuanced analysis of firms' practices and performance than is possible with existing frameworks. It supports the essential work of monitoring whether or not individual firms are adhering to regulations. However, it is equally beneficial in terms enabling institutional investors to make fuller assessments of the likely risks associated with investing in particular firms, so it can therefore also be seen as providing key support to their due diligence efforts. This is an especially valuable contribution, since firms that are subject to a higher level of external scrutiny may engage in selective disclosure (Marquis & Toffel, 2016). A framework focused on measuring quality could help identify parts of such behavior after validation, thereby further highlighting its importance.

In terms of ensuring the bilingual functioning of the framework, using NLP for Arabic-language documents created several practical difficulties. For example, extracting PDF text written from right to left

proved challenging. Similarly, difficulties were posed by the presence in the source documents of morphologically complex word forms that required normalization. The mixed-language nature of many of the reports also required close attention, since it is common among some firms to produce reports that contain passages in both English and Arabic. Integrating AraBERT (Antoun et al., 2020) and Farasa normalization via CAMeL Tools (Obeid et al., 2020) assisted considerably with overcoming the language challenges. However, working with two such different languages was more complex than, say, creating a framework that could operate in countries with several official languages written in the same script or with left to right text. These challenges are not specific to Saudi Arabia; however, their interaction with Arabic–English reporting makes them especially relevant for this corpus.

Alongside the bilingual language setting, Saudi Arabia's evolving regulatory landscape introduced another challenging but interesting layer of complexity to the study. The country's drive to improve transparency as part of its Vision 2030 reforms has created a "halfway house" situation, in effect an interim period in which different firms are taking different approaches to disclosure or are moving at different speeds towards full compliance with the new regulations. Far from this rendering Saudi Arabia an unsuitable context for the research, it makes it an ideal setting, since the framework has to be capable of functioning in countries where regulations and practices are evolving. Indeed, it can be argued that investors in these contexts have a greater need than others to be able to gauge the quality of firms' ESG disclosure. Nevertheless, the changes under way in Saudi Arabia did contribute to the study's challenges, and the requirement for enhanced sustainability disclosures (Capital Market Authority Saudi Arabia, 2021) created a natural regulatory boundary during the study period. If a validated ESG_C dataset were established in future research, this would enable examination of patterns before and after the regulatory changes, which would be interesting to analyze.

Finally, the study's selection bias towards large-cap firms was a conscious design choice to exclude firms that may not have had an adequate level of IR infrastructure to meet the study's purposes. It also ensured the sample reflected the state of the Saudi ESG disclosure ecosystem. This is because the firms most capable of producing complete, machine-readable annual reports are those over the SAR 500 million threshold set for inclusion in the study sample. However, this decision left open the question of whether the framework would perform well on smaller-cap company reports. As a result, the decision acted as a clear methodological boundary while also causing selection bias.

6 Conclusions

This paper presents a methodological framework for an NLP-based bilingual proposed system designed to measure ESG disclosure quality in the Saudi market, together with a corpus of 250 annual reports from 50 Tadawul-listed companies for the period 2019–2023. It creates an Arabic–English NLP pipeline and a four-dimensional ESG composite quality index (ESG_C), grounded in signaling theory, stakeholder theory, hard disclosure theory, and legitimacy theory. The framework was not tested empirically and no statistical inferences are drawn. The sample covers 21.6% of Tadawul-listed companies and 13 of 21 industry groups represented on the index. It has an acknowledged large-cap orientation.

6.1 Contributions

This research makes several contributions. Firstly, it offers one of the first documented attempts to design a framework for measuring the quality of ESG disclosures by Saudi listed firms, by means of a bilingual NLP-based system that can be used for both Arabic- and English-language corporate annual reports. Secondly, the study compiles a fully attributed corpus of 250 annual reports with company-level, sector-level, and source-level documentation. Because annual reports are copyrighted, the corpus itself is not released in this paper. However, the company list, source categories, dictionary, code, and derived non-copyright metadata should be made available in future implementations where permissions allow. Finally, it offers a rigorous sample transparency report characterizing sector coverage, selection bias, and the exclusion rationale.

6.2 Limitations and Future Research Directions

Despite its significant contributions, this paper has a number of limitations that should be acknowledged. Firstly, it presents a framework and corpus, not validated empirical findings, and therefore does not present evidence on ESG disclosure quality. The F1 = 0.87 metric applies to the sentence-level binary ESG classification component on a 300-sentence hold-out set only. As noted in Section 3.3, validation was limited

to sentence-level classification of the report contents. Future researchers could address this issue by achieving a full corpus-level validation of the ESG_C framework. There are several ways this could be achieved. Methods could include large-scale annotation by humans and inter-rater reliability testing. Other viable options such as component sensitivity analysis, alternative weighting calibration, and empirical benchmarking against established ESG disclosure datasets should also be considered.

As discussed in detail in Section 4, there are further limitations relating to the sample scope and representativeness. For example, the sample covers 21.6% of Tadawul-listed companies by count and 13 of 21 industry groups, meaning eight industry groups are absent. As a result, the findings cannot be generalized to other firms on the Tadawul. It is also acknowledged that the SAR 500 million market cap threshold introduces a size bias, meaning smaller firms are overlooked. These limitations could be remedied in future studies by revising the inclusion criteria to draw in a higher proportion of firms and sectors and to ensure small- and mid-cap firms are also represented in the sample.

The quality of the OCR represents a further limitation. Approximately 12% of the company reports required OCR processing. The estimated character-level accuracy of 94.3% implies a ~5.7% error rate, which could affect the NLP classification. Another issue is that ESG disclosures that were made only in tables, figures, and footnotes may have been overlooked, as the pipeline treated them as complementary to the text rather than as new or additional material.

In addition, the ESG dictionary, while containing 300 terms, had incomplete coverage of sector-specific Arabic terms, and the 1,200-sentence AraBERT sets could be considered insufficient. Future validation requires a larger dataset with multiple annotators, diverse sectors, and clear adjudication rules. Future research should also expand the dictionary, cover more sector-specific Arabic ESG terminology, and expand training the datasets to improve classification across Saudi and GCC reporting. Regarding the 82 firms excluded from the study owing to missing reports, inaccessible files, or non-processable PDF formats, it would be preferable if future studies systematically logged failed access attempts. Future researchers could also validate the findings by comparing them with third-party ESG quality assessments or conducting longitudinal testing of regulatory events. Comparing the results with findings from other GCC states would enable cross-country comparisons in similar markets. It would also be valuable to conduct full ground-truth annotation of a representative corpus subset to validate the ESG_C scoring. In addition, empirical estimation of the proposed determinants model could also be conducted. These latter two avenues could only be explored after the pipeline outputs have been validated.

Future empirical implementations should include the code, preprocessing scripts, ESG dictionary, model configuration, validation partitions, exclusion logs, and derived firm-year outputs, while ensuring that copyrighted reports are not redistributed. As acknowledged in Section 3.5, it is also important to evaluate the model using firm-clustered standard errors, heteroskedasticity-robust inference, alternative fixed effect structures, and wider control sets. Validation of the weighting sensitivity should also be considered in future research (see Section 3.3). Finally, as noted in Section 3.2, future researchers could categorize the reports based on language (Arabic only, English only, mixed) and record retrieval dates and store archive checksums to improve reproducibility, since the possibility of version differences in the documents cannot be excluded given the different sources.

7 References

Abdullah, A., Talha, M., & Hadi, N. U. (2025). The Effects of ESG Performance on Financial Performance: Evidence From GCC Countries. *Corporate Social Responsibility and Environmental Management.* [DOI]

Alotaibi, K.O.; Hussainey, K. (2016). Determinants of CSR disclosure quantity and quality: Evidence from non-financial listed firms in Saudi Arabia. *International Journal of Disclosure and Governance,* 13, 364-393. [DOI]

Antoun, W., Baly, F., Hajj, H. (2020). *AraBERT: Transformer-based model for Arabic language understanding* [Conference paper]. 4th Workshop on Open-Source Arabic Corpora and Processing Tools, LREC 2020, Marseille, France. https://aclanthology.org/2020.osact-1.2/

Argaam. Financial Reports Archive. Available online: https://www.argaam.com (accessed on 15 January 2026).

Bamahros, H. M, Alquhaif, A, Qasem, A, Wan-Hussin, WN, Thomran, M, Al-Duais, SD, Shukeri SN, & Khojally, H. M. A. (2022). Corporate governance mechanisms and ESG reporting: Evidence from the Saudi stock market. *Sustainability,* 14(10), 6202. [DOI]

Berg, F., Kölbel, J.F. & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. Review of Finance, 26, 1315–1344. [DOI]

Bingler, J.A.; Kraus, M.; Leippold, M.; & Webersinke, N. (2022). Cheap talk and cherry picking: What ClimateBert has to say on corporate climate risk disclosures. Finance Research Letters, 47, 102776. [DOI]

Capital Market Authority Saudi Arabia. (2021). *Sustainability and ESG Reporting Guidelines for Issuers of Securities Listed on the Stock Exchange*; CMA: Riyadh, Saudi Arabia, 2021. Available online: https://www.saudiexchange.sa/wps/portal/saudiexchange/listing/issuer-guides/esgguidelines

Cho, C.H.; Guidry, R.P.; Hageman, A.M. & Patten, D. M. (2012). Do actions speak louder than words? An empirical investigation of corporate environmental reputation. *Accounting, Organizations and Society*, 37, 14–25. [DOI]

Clarkson, P.M.; Li, Y.; Richardson, G.D. & Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure. *Accounting, Organizations and Society*, 33, 303–327. [DOI]

Deegan, C. (2002). The legitimising effect of social and environmental disclosures: A theoretical foundation. *Accounting, Auditing & Accountability Journal*, 15, 282–311. [DOI]

Dhaliwal, D.; Li, O.Z.; Tsang, A.; & Yang, Y. G. (2011). Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *The Accounting Review*, 86, 59–100. [DOI]

Freeman, R.E. (1984). *Strategic Management: A Stakeholder Approach*. Pitman Publishing: Boston, MA, USA, 1984.

Global Reporting Initiative (2021). *GRI Universal Standards 2021*; GRI Secretariat: Amsterdam, The Netherlands, 2021.

Gray, R., Kouhy, R.; & Lavers, S. (1995). Corporate social and environmental reporting: A review of the literature and a longitudinal study of UK disclosure. *Account. Accounting, Auditing & Accountability Journal*, 1995, 8, 47–77. [DOI]

Habbash, M. (2016). Corporate governance and corporate social responsibility disclosure: Evidence from Saudi Arabia. Soc. *Social Responsibility Journal*, 12, 740–754. [DOI]

Haniffa, R. M, & Cooke, T. E. (2005). The impact of culture and governance on corporate social reporting. *Journal of Accounting and Public Policy*, 24, 391–430. [DOI]

Hummel, K., & Schlick, C. (2016). The relationship between sustainability performance and sustainability disclosure: Reconciling voluntary disclosure theory and legitimacy theory. *Journal of Accounting and Public Policy*, 35, 455–476. [DOI]

Ioannou, I., & Serafeim, G. (2019). *The consequences of mandatory corporate sustainability reporting*. In A. McWilliams, D. E. Rupp, D. S. Siegel, G. K. Stahl, & D. A. Waldman (Eds.), The Oxford handbook of corporate social responsibility: Psychological and organizational perspectives (Online ed.). Oxford Academic. [DOI]

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54, 1187–1230. [DOI]

Marquis, C., & Toffel, M. W. (2016). Scrutiny, norms, and selective disclosure: A global study of greenwashing. *Organization Science*, 27, 483–504. [DOI]

Michelon, G.; Pilonato, S., & Ricceri, F. (2015). *CSR reporting practices and the quality of disclosure: An empirical analysis*. Critical Perspectives on Accounting, 33, 59–78. [DOI]

Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A., & Habash, N. (2020). CAMeL tools: An open source Python toolkit for Arabic natural language processing. In Proceedings of the Twelfth Language Resources and Evaluation Conference (pp. 7022–7032). European Language Resources Association. https://aclanthology.org/2020.lrec-1.868/

Saudi Exchange (Tadawul). *Listed Companies Database*. Available online: https://www.saudiexchange.sa (accessed on 15 January 2026).

Spence, M. (1973). Job market signaling. *Q. J. Econ.* 1973, 87, 355–374. [DOI]

Suchman, M.C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20, 571–610. [DOI]

Sustainability Accounting Standards Board. (2018). *SASB Standards Overview: Sector and Industry Standards; SASB: San Francisco*, CA, USA, 2018.

Task Force on Climate-related Financial Disclosures. (2021). *Implementing the Recommendations of the Task Force on Climate-related Financial Disclosures*; TCFD Secretariat: Basel, Switzerland, 2021.

United Nations Principles for Responsible Investment. (2022). *Annual Report 2022*; UNPRI: London, UK, 2022.Available online: https://dwtyzx6upklss.cloudfront.net/Uploads/b/f/m/pri_annual_report_2022_689047.pdf (accessed on 15 January 2026).