

Article

Detecting Greenwashing in ESG Disclosure: An NLP-Based Analysis of Central and Eastern European Firms

Adriana AnaMaria Davidescu ^{1,2}, Eduard Mihai Manta ¹, Ioana Bîrlan ^{3,*}, Alexandra-Mădălina Miler ⁴ and Sorin-Cristian Niță ⁵

¹ Department of Statistics and Econometrics, The Bucharest University of Economic Studies, 010552 Bucharest, Romania; adriana.alexandru@csie.ase.ro or adriana.davidescu@incsmpls.ro (A.A.D.); eduard.manta@csie.ase.ro (E.M.M.)

² Department of Education, Training and Labour Market, National Scientific Research Institute for Labour and Social Protection, 010643 Bucharest, Romania

³ Doctoral School of Economic Cybernetics and Statistics, The Bucharest University of Economic Studies, 010552 Bucharest, Romania

⁴ Faculty of Cybernetics, Statistics and Economic Informatics, The Bucharest University of Economic Studies, 010552 Bucharest, Romania; mileralexandra19@stud.ase.ro

⁵ Department of UNESCO Chair for Business Administration, Faculty of Business Administration, The Bucharest University of Economic Studies, 010552 Bucharest, Romania; cristians.nita@fabiz.ase.ro

* Correspondence: ioana.birlan@csie.ase.ro

Abstract

The rapid expansion of corporate sustainability reporting has increased transparency requirements while raising concerns about greenwashing driven by selective, narrative-based disclosure. This study assesses the credibility of Environmental, Social, and Governance (ESG) communication by comparing corporate sustainability reports with external media coverage for a sample of 204 large firms operating in Central and Eastern Europe in 2023. Using natural language processing techniques, the analysis constructs a Greenwashing Severity Index (GSI) that captures discrepancies between firms' ESG self-representation and external public narratives. The index combines ESG-specific focus measures, sentiment analysis, TF-IDF-based term weighting, and topic modeling to quantify imbalances in ESG communication. Results indicate moderate but widespread greenwashing across countries, industries, and firm sizes, with substantial heterogeneity linked to differences in regulatory maturity and stakeholder scrutiny. Higher alignment between corporate disclosures and external narratives is observed among larger firms and in sectors subject to stronger public accountability, while finance, aviation, and online commerce exhibit higher greenwashing severity. A propensity score matching analysis further shows that firms with imbalanced emphasis across ESG dimensions display significantly higher GSI values, consistent with strategic disclosure behavior rather than substantive sustainability engagement. Overall, the findings demonstrate that transparency alone is insufficient to ensure credible ESG communication, highlighting the need for EU sustainability governance to move beyond disclosure-based compliance toward digitalized, data-driven monitoring frameworks that systematically integrate external information sources to curb strategic ESG misrepresentation and enhance corporate accountability under evolving regulatory regimes.

Keywords: big data analytics; corporate transparency; digitalization; ESG disclosure; greenwashing; natural language processing; sustainability reporting

Academic Editors: Wen-Hsien Tsai

Received: 20 December 2025

Revised: 21 January 2026

Accepted: 22 January 2026

Published: 2 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

The rapid expansion of corporate sustainability reporting has generated vast amounts of unstructured textual information, creating new challenges for transparency, comparability, and effective governance. While Environmental, Social, and Governance (ESG) disclosures are intended to support sustainable business practices and inform regulatory oversight and investment decisions, their increasing volume and narrative complexity have also heightened the risk of selective reporting and symbolic compliance. Recent empirical evidence indicates that extensive sustainability reporting may improve firms' apparent sustainability performance without necessarily translating into real environmental or social improvements, thereby amplifying greenwashing risk rather than mitigating it [1]. Within this context, greenwashing has emerged as a critical governance-related issue, capturing the divergence between corporate sustainability claims and observed practices.

Originally introduced in 1986 to describe environmentally misleading communication, the concept has since evolved into a multidimensional concern encompassing environmental, social, and governance dimensions. Contemporary conceptual work increasingly defines greenwashing as the gap between "apparent" sustainability performance communicated through reporting and "real" sustainability performance reflected in underlying practices [2,3]. However, as highlighted in [4], the absence of conceptual and methodological consensus continues to constrain the ability of regulators and stakeholders to systematically identify and assess deceptive sustainability communication. This underscores the need for structured evidence-based method approaches to evaluating ESG transparency. In this context, Natural Language Processing (NLP) provides a core set of tools for operationalizing such approaches, enabling scalable, reproducible analysis of sustainability reporting narratives and their alignment with external information sources.

The increasing relevance of greenwashing reflects broader challenges in the governance of corporate sustainability. While the concept initially emerged in relation to environmentally misleading communication, subsequent research has shown that greenwashing also encompasses distortions in social responsibility and corporate governance practices. As noted in [5], heightened awareness of the economic, social, and institutional consequences of corporate activity has intensified demands for transparency from a wide range of stakeholders, including investors, consumers, financial institutions, and public authorities. These developments have reinforced the role of the Environmental, Social, and Governance (ESG) framework as a central reference for evaluating corporate sustainability. Within this framework, sustainability performance is assessed through three interconnected dimensions: environmental impact and resource use, social relations with employees and communities, and governance practices related to accountability, transparency, and ethical conduct. From this perspective, greenwashing represents an ESG-wide governance concern rather than a narrowly environmental issue.

At the same time, the nature of sustainability reporting has changed substantially. Corporate ESG disclosures increasingly rely on narrative and qualitative content, resulting in large volumes of textual material that are difficult to assess using traditional, manual approaches. As reporting requirements expand and sustainability communication becomes more detailed, ensuring transparency and comparability requires analytical methods capable of systematically processing and evaluating complex textual information. In this context, Natural Language Processing (NLP) has emerged as a core digital tool for analyzing sustainability reporting at scale by enabling systematic, replicable comparisons between corporate narratives and independently produced information [6]. Digitalization, understood as the use of structured, computer-assisted, and data-driven techniques for analyzing information, has therefore become an important enabling factor in sustainability assessment and governance oversight.

Within this framework, the present study examines the incidence and characteristics of greenwashing among corporations operating in Central and Eastern Europe (CEE). The empirical analysis focuses on sustainability reporting and related media coverage for the year 2023, capturing both corporate self-representation and external perceptions. Differences in language, tone, and thematic emphasis are analyzed using Natural Language Processing (NLP) techniques. On this basis, a Greenwashing Severity Index (GSI) is constructed to quantify the divergence between corporate sustainability reporting and external media narratives. The sample comprises 204 firms listed in the Coface Top 500 for Central and Eastern Europe, covering multiple industries and firm sizes. The regional focus is analytically relevant given the heterogeneity of governance structures, regulatory capacity, and sustainability reporting practices across CEE countries.

The objective of the study is to identify, measure, and explain variation in greenwashing across institutional and organizational contexts. To this end, the analysis evaluates the consistency and credibility of sustainability reporting narratives across environmental, social, and governance dimensions by combining sentiment analysis, term-weighting techniques, and topic modeling. The systematic comparison between corporate sustainability reporting and independent media coverage provides an empirical basis for assessing transparency as narrative alignment rather than disclosure volume alone, illustrating how digitally supported, data-driven methods can enhance governance-oriented evaluation of sustainability communication.

Consistent with the perspective outlined in [6], greenwashing is conceptualized in this study as an ESG-wide phenomenon rather than a narrowly environmental issue. The ESG framework therefore serves as the conceptual foundation of the analysis, enabling an integrated examination of corporate sustainability practices. Within this framework, transparency is understood not simply as the quantity of information disclosed, but as an outcome of governance quality, stakeholder scrutiny, and monitoring processes that increasingly rely on structured analytical tools.

The paper seeks to answer the following policy-relevant questions:

1. Which countries in Central and Eastern Europe exhibit the highest levels of greenwashing, and how do these patterns relate to differences in governance quality and regulatory enforcement?
2. Which industries are most exposed to greenwashing risks, and what sector-specific characteristics help explain these patterns?
3. How does firm size influence ESG transparency and the credibility of sustainability communication?
4. Which countries and sectors display stronger alignment between corporate sustainability reporting and external narratives, indicating more effective accountability mechanisms?

Building on this research design, the hypotheses operationalize the proposed relationships between institutional context, firm characteristics, and observed greenwashing severity. The first hypothesis posits that the degree of greenwashing varies systematically across countries and industries due to differences in regulatory maturity, stakeholder pressure, and disclosure norms. The second hypothesis specifies that firm size is a determinant of sustainability reporting transparency, operationalized as lower Greenwashing Severity Index (GSI) values, indicating higher alignment between corporate sustainability reporting and external media narratives. Larger companies are therefore expected to exhibit lower levels of greenwashing due to greater public scrutiny, compliance obligations, and reputational exposure.

This study enhances the literature on corporate sustainability communication and greenwashing detection by proposing a new NLP-based Greenwashing Severity Index (GSI) which measures the difference between corporate ESG disclosures and external

media narratives. Unlike previous studies which have mainly focused on qualitative approaches or one-dimensional environmental variables, this study proposes a multi-dimensional, data-driven metric which focuses on environmental, social and governance aspects. By combining sentiment analysis, a TF-IDF term weight analysis and topic modeling (LDA), a complete analytical framework is created which measures both the tone and the thematic variation in the sustainability communications. By focusing on Central and Eastern Europe, the study addresses a significant geographical gap in the literature, which has predominantly examined Western economies, and provides policy-relevant insights for regulators, investors, and public authorities seeking to strengthen accountability in sustainability reporting.

The paper is structured in two main parts: one theoretical and one practical. The theoretical part reviews some of the most relevant studies on the greenwashing phenomenon, the importance of ESG criteria, and the integration of natural language processing to assess the phenomenon. The practical part is organized into several sections. The first part presents the data used in the analysis, followed by a brief description of the research methodology. The final section highlights the research results, which represent the core component of the paper. Lastly, the conclusions regarding the greenwashing phenomenon in Central and Eastern Europe are presented.

1.1. Identifying the Greenwashing Phenomenon

As environmental awareness expands, companies increasingly rely on green marketing to promote their sustainability credentials. Yet, as ref. [7] conceptually argues, this practice often risks degenerating into greenwashing, where environmental claims are strategically exaggerated or misleading. Such distortions may appear subtle but carry significant implications for consumer trust and market integrity. From a governance perspective, greenwashing reflects a structural asymmetry between firms' incentives to manage external perceptions and stakeholders' limited ability to verify underlying sustainability practices. Transparency, verifiability, and accountability therefore emerge as necessary conditions for ensuring that sustainability reporting functions as a mechanism of credible disclosure rather than symbolic communication.

Recent research has expanded the discussion on greenwashing beyond corporate ethics, examining its wider implications for environmental sustainability, social welfare, and governance. According to ref. [8], greenwashing distorts market dynamics, misallocates resources, and undermines regulatory trust. The environmental consequences of deceptive sustainability claims—such as deforestation, pollution, and resource depletion—are accompanied by social effects, including misinformation and declining confidence in corporate reporting. These outcomes show that greenwashing is not merely a marketing issue but a structural challenge to sustainable development.

A first stream of the literature focuses on stakeholder interpretation and perception mechanisms. Early studies [9–12] demonstrate that vague, unverifiable, or ambiguous sustainability claims increase skepticism and reduce trust, particularly among environmentally aware stakeholders. Rather than viewing these contributions in isolation, the literature converges on a common mechanism: low-quality sustainability reporting weakens credibility precisely where transparency is expected to enhance trust. This insight helps explain why disclosure quantity alone is insufficient to mitigate greenwashing risks.

The reputational and behavioral consequences of misleading sustainability reporting are further documented in experimental and survey-based research. Studies show that perceived greenwashing leads to brand avoidance, negative word-of-mouth, and declining purchase intention [13–15]. Within the Stimulus–Organism–Response (SOR) framework, sustainability disclosures act as the stimulus, stakeholder credibility assessments represent the organismic state, and behavioral reactions such as distrust or avoidance

constitute the response. This framework clarifies how narrative misalignment in sustainability reporting translates into tangible economic and reputational costs.

Efforts to understand and mitigate greenwashing are increasingly shifting from descriptive assessments toward systemic, governance-oriented, and data-driven approaches. Evidence from [16] demonstrates that institutional digitalization can materially constrain greenwashing by strengthening monitoring capacity and regulatory oversight. However, such institutional mechanisms primarily address incentives and enforcement conditions rather than providing direct, firm-level measures of narrative misalignment, limiting their usefulness for cross-firm and cross-sector comparability.

At the same time, persistent conceptual and methodological challenges continue to hinder empirical progress. As emphasized in [17], the absence of standardized definitions and measurement frameworks has resulted in fragmented evidence, often relying on perception-based indicators or hypothetical scenarios. These approaches are informative about stakeholder responses but insufficient for identifying greenwashing as an observable discrepancy between reported and underlying practices, thereby weakening their policy and governance relevance.

Empirical attempts to operationalize greenwashing therefore yield mixed and context-dependent results. While ref. [18] finds limited evidence of systematic greenwashing in U.S. metals firms, its conclusions are constrained by sector specificity and reliance on self-reported disclosures. This highlights a broader limitation of approaches that assess sustainability reporting in isolation, without benchmarking it against independent external information sources. In contrast, ref. [19] demonstrates that artificial intelligence and NLP techniques can systematically detect inconsistencies between corporate sustainability narratives and observed behavior. Yet even advanced models such as BERT-based classifiers often focus on environmental dimensions or binary classification, leaving open the question of how to capture the severity, multidimensionality, and governance relevance of greenwashing.

Taken together, this literature suggests that while institutional digitalization and AI-based tools each contribute to mitigating greenwashing, neither alone fully resolves the challenge of measuring ESG-wide credibility gaps in sustainability reporting. This unresolved tension motivates the need for discrepancy-based, multidimensional indicators such as the Greenwashing Severity Index proposed in this study that explicitly quantify narrative misalignment between corporate sustainability reporting and independent external narratives.

1.2. Integration of ESG Criteria in the Assessment of the Greenwashing Phenomenon

A recurring insight across the ESG literature is the tension between formal compliance with sustainability reporting requirements and the substantive integration of ESG principles into corporate strategy and operations. Firms may comply with increasingly detailed disclosure frameworks while engaging in non-strategic ESG activities that primarily serve reputational or legitimacy purposes rather than reflecting genuine sustainability commitments. This pattern, often described as formalized transparency, provides a useful lens for understanding greenwashing as a governance-related issue rather than merely a communication failure.

The concept of Corporate Social Responsibility (CSR) is closely intertwined with the study of greenwashing, as both reflect the tension between corporate ethical commitments and their public representation. As demonstrated by [20], strong CSR performance enhances firms' access to capital through reduced agency costs and lower information asymmetry. However, subsequent evidence suggests that transparency achieved through sustainability reporting does not necessarily translate into substantive ESG performance, particularly when disclosure incentives are weakly aligned with operational change.

The Non-Financial Reporting Directive 2014/95/EU marked a significant milestone in expanding disclosure obligations for large European companies, requiring detailed reporting on environmental, social, and governance aspects. Ref. [21], in a comparative study between the United Kingdom and Italy, shows that the quality of non-financial information (NFI) differs substantially across national contexts, with stronger regulatory environments yielding higher compliance and disclosure quality. These findings emphasize the role of national governance and institutional enforcement in shaping the integrity of corporate sustainability reporting.

With the global diffusion of ESG standards, assessing firms' performance through environmental, social, and governance criteria has become a cornerstone of responsible business practice. ESG performance evaluates concrete actions, outcomes, and governance mechanisms that reflect a company's genuine commitment to sustainability [22]. Yet, as ref. [23] notes in its analysis of Swedish firms, improvements in the quality and quantity of ESG reporting have not necessarily translated into better ESG performance. The divergence between reporting and actual outcomes highlights a growing concern that disclosure frameworks may incentivize form over substance. Strengthening the link between ESG reporting and measurable sustainability outcomes remains a pressing challenge.

The interaction between corporate governance and greenwashing behavior is further explored in emerging markets. Using a large sample of Chinese A-share listed firms, ref. [24] finds that greater executive compensation gaps amplify greenwashing by increasing corporate risk-taking and weakening ethical oversight. This effect is more pronounced in firms with weak environmental awareness, high earnings management, and limited media scrutiny, illustrating how internal governance dynamics can shape sustainability credibility. Similarly, ref. [25] identifies artificial intelligence (AI) as a moderating factor that can mitigate greenwashing behavior by improving the quality of ESG disclosures. AI enhances transparency and accountability, particularly within state-owned enterprises and industries subject to stringent environmental regulations, where it lowers financial constraints and management costs while promoting green innovation.

A key concern in the assessment of ESG practices is the persistence of formal compliance without substantive integration, particularly in institutional investment contexts. In this respect, ref. [26] shows that ESG integration in investment decision-making across Central and Eastern Europe remains largely compliance-driven rather than strategically embedded. Its analysis of Romanian pension funds indicates that while frameworks such as the SFDR, EU Taxonomy, and CSRD have improved disclosure and formal accountability, they have not ensured the genuine incorporation of sustainability principles into investment behavior. This pattern illustrates how ESG reporting may function primarily as reputational signaling rather than as evidence of structural change, reflecting a regional manifestation of greenwashing. The findings reinforce a central argument of this study: transparency requirements alone cannot ensure authenticity, as the real challenge lies in the interpretive quality and alignment of ESG data with verifiable sustainability outcomes.

Recent studies further highlight the nuanced relationship between ESG disclosure and firm performance. Ref. [27], through a meta-analysis of 80 studies, finds a small but significant positive link between ESG disclosure and accounting-based performance, while market-based effects remain weak or inconsistent. This suggests that transparency may strengthen internal efficiency and stakeholder trust without necessarily being priced by investors. In contrast, ref. [28] shows that ESG-oriented technology firms in Japan outperform both market-wide and non-ESG technology benchmarks, indicating that when ESG disclosure reflects substantive strategic integration rather than symbolic compliance, it can support long-term competitiveness and value creation. Taken together, these findings indicate that the impact of ESG disclosure depends on its depth and strategic embedding.

At a global level, the synthesis by [29] provides a comprehensive overview of greenwashing research, identifying three dominant themes in sustainability reporting: exaggeration of environmental efforts, the gap between ESG disclosure and performance, and legitimacy-driven communication strategies. The study highlights that institutional and cultural differences significantly shape the manifestation of greenwashing between G7 and non-G7 countries, emphasizing the need for greater attention to context-specific governance mechanisms. It also reinforces the role of regulation and digital tools as essential mechanisms for detecting and deterring deceptive ESG reporting.

The relationship between ESG indicators and financial performance remains an area of active debate. While refs. [30,31] demonstrate that ESG indicators guide socially responsible investment decisions, ref. [32] finds that in the Serbian energy sector, ESG indicators often show no direct correlation with financial performance. This suggests that the integration of ESG into business strategy remains uneven, with firms selectively reporting indicators without embedding them in core operations.

The methodological innovations of [33] mark a significant step toward integrating ESG criteria into greenwashing assessment. Using a diverse dataset of 702 companies, the study applies advanced natural language processing (NLP) and an ESG-based Greenwashing Severity Index (GSI) to detect discrepancies in sustainability reporting. The results reveal substantial variation across sectors and countries, with smaller firms exhibiting lower levels of deceptive reporting and Portugal emerging as a high-risk case in the social dimension. These findings underline the need for transparent ESG reporting and third-party verification to strengthen public confidence and regulatory oversight.

Advancements in machine learning have further expanded the analytical toolkit for assessing corporate transparency. Ref. [34] develops NLP models trained on over 13.8 million corporate and media texts, enabling automated classification of ESG-related content across environmental, social, and governance dimensions. This approach enhances the precision of sustainability assessments and contributes to bridging the gap between narrative disclosures and measurable performance. However, the authors note that such models may face generalization challenges and require refinement to capture nuanced communicative strategies beyond textual data.

Ref. [35] extends this discussion by comparing internal ESG sentiment with public opinion extracted from social media for pharmaceutical companies. The lack of correlation between internal and external sentiment indicates potential “green disinformation,” suggesting that companies may use ESG narratives as reputation management tools rather than reflections of substantive action.

The literature reviewed reveals a consistent tension between the form and substance of corporate sustainability communication. Across diverse institutional contexts, studies converge on a critical insight: transparency in ESG disclosure, while necessary, is not a sufficient indicator of authenticity. Firms may comply with reporting frameworks yet fail to embed sustainability principles within their strategic and operational structures, resulting in what can be termed formalized transparency—the appearance of accountability without corresponding behavioral transformation. This paradox underscores the complexity of assessing greenwashing empirically, as the boundary between strategic communication and misrepresentation often lies in the tone, context, and framing of sustainability narratives.

In this light, the present study adopts a data-driven approach to quantify such discrepancies by examining the divergence between corporate self-representation and independent media perceptions. By constructing a Greenwashing Severity Index (GSI) through Natural Language Processing (NLP) techniques, the analysis moves beyond declarative reporting toward an empirical evaluation of credibility and consistency in corporate sustainability communication. Relative to existing approaches, this framework

contributes by operationalizing greenwashing as an observable, multidimensional discrepancy between reported ESG narratives and externally generated signals, rather than as a perception-based or purely environmental construct.

More importantly, the contribution of this study extends beyond measurement. By systematically identifying where and how sustainability reporting diverges from external narratives, the GSI provides an empirical basis for distinguishing symbolic compliance from substantive ESG engagement. In doing so, the analysis supports efforts to move sustainability reporting from a formal disclosure exercise toward a governance mechanism capable of incentivizing genuine ESG integration into corporate development strategies. This methodological framework directly builds on the theoretical foundations established in the literature, linking the abstract concept of greenwashing with measurable linguistic and semantic patterns that reflect the authenticity—or absence—of corporate ESG commitment.

2. Materials and Methods

This study employs a multi-stage analytical design that integrates large-scale document collection, lexicon-based linguistic analysis, and topic modeling techniques to evaluate the credibility of corporate sustainability narratives in Central and Eastern Europe (CEE). The methodological framework builds on the NLP-based greenwashing detection model proposed by [28], adapting it to a regional and comparative context. Specifically, the approach captures discrepancies between firms' self-reported sustainability performance and how these claims are reflected in external media discourse. The analysis focuses on the year 2023, selected as the most recent period for which comprehensive and comparable sustainability reports were publicly available across the CEE region. This timeframe provides the broadest possible coverage of corporate disclosures, coinciding with intensified ESG reporting obligations under evolving EU regulations and growing public scrutiny of firms' environmental and social practices.

The construction of the four dictionaries followed a hybrid and iterative procedure. First, an initial pool of candidate terms was compiled from existing ESG-related lexicons and prior studies on sustainability reporting, greenwashing detection, and ESG text analytics. These sources included academic NLP applications in sustainability research, ESG disclosure frameworks, and regulatory terminology commonly used in corporate non-financial reporting.

In a second step, the preliminary lists were manually curated by the authors to ensure semantic relevance, contextual clarity, and alignment with the objectives of greenwashing detection. Terms that were excessively generic, context-dependent, or prone to misclassification were excluded. The remaining terms were assigned to one of the four dictionaries—Greenwashing, Environmental, Social, or Governance—based on their dominant semantic meaning.

Validation was conducted through iterative testing on a pilot corpus of sustainability reports and corresponding media articles. Term frequencies and contextual usage were examined to verify consistency across sectors and countries. Only terms exhibiting stable and meaningful ESG-related usage were retained in the final dictionaries, which are reported in Appendixes A.1–A.4.

2.1. Data Sources and Sampling

The dataset was constructed from publicly accessible corporate sustainability or annual reports and news coverage pertaining to the same firms. The starting point for the sampling procedure was the Coface Top 500 CEE (2023) ranking [36], which lists the largest companies across Central and Eastern Europe. This database offers extensive coverage of industries ranging from agriculture, energy, and manufacturing to technology, healthcare, aviation, finance, and consumer goods.

In addition, to support qualitative data analysis, four dictionaries were created that were used to identify and evaluate greenwashing statements and behaviors in company reports and public communications. The dictionaries were created in English, due to the fact that both company reports and news about them are in English. The dictionaries were designed to capture linguistic nuances across environmental, social, and governance dimensions and are documented in Appendixes A.1–A.4.

The Greenwashing dictionary (A1) contains 1140 positive words, marked with the value -1 , which would normally be used in the descriptions of companies that are not suspected of false sustainable practices, and 990 negative words, marked with the value 1 , which would induce a reader's susceptibility. For example, marked with a value of -1 are words such as "sustainability", "climate-friendly", "ethical" or "low-impact". At the opposite pole, words such as "pollution", "irresponsible", "anti-environmental" or "eco-fraud" were marked with a value of 1 .

The three ESG dictionaries, the environmental dictionary (A2), the social dictionary (A3) and the governance dictionary (A4) each contain 122 words taken from the lexicon of each field. They are a set of standards for measuring a business's impact on society and the environment, as well as its transparency and accountability.

The Environmental Dictionary (E) brings together key terms related to environmental sustainability, helping to guide responsible decision-making and encourage actions that preserve natural resources for the future. It includes expressions such as plastic reduction, water conservation, and recycling.

The Social Dictionary (S) contains terms that describe the social aspects of a company's operations and their broader effects on society. It covers topics such as labor standards, human rights, diversity and inclusion, community involvement, and customer well-being. Examples of terms include civic responsibility, ethical recruiting, and health and safety.

The Governance Dictionary (G) focuses on the systems and principles that shape how a company is managed and held accountable. It includes concepts related to organizational structure, internal controls, transparency, and shareholder rights, with examples such as anti-corruption, code of ethics, and voting rights.

The collection and processing of textual data followed several steps designed to ensure reliability and comparability across sources. News articles were retrieved using the Python 3.11 package *gnews* [37], which aggregates content from established international media outlets. To obtain full article texts, customized web-scraping procedures were subsequently applied. Relevance filtering was conducted in two stages. First, articles were retained only if the company name appeared in the headline or within the main body of the text, ensuring a direct link between the content and the firm under analysis. Second, keyword-based screening was applied to exclude items unrelated to sustainability, ESG practices, or corporate responsibility. Duplicate articles, syndicated copies, and near-identical news items were removed using textual similarity checks. This process resulted in a refined corpus covering approximately 320 companies for which sufficiently rich and relevant media text was available. Sustainability and annual reports were collected directly from official corporate websites or investor relations sections. Both corporate reports and media texts were subjected to identical preprocessing steps, including lowercasing, punctuation normalization, tokenization, lemmatization, and the removal of stopwords and non-informative symbols. This harmonized preprocessing ensured that corporate disclosures and external narratives could be compared on a consistent linguistic basis.

2.2. Sentiment Analysis and Comparative Scoring

Sentiment analysis, a core natural language processing (NLP) technique, was applied to detect emotional tone and assess the authenticity of sustainability communication. Each word in the Greenwashing Dictionary (A1) was assigned a sentiment score ranging from -3 to $+3$, where -3 denotes ecological sincerity and $+3$ represents a strong likelihood of greenwashing. The sentiment calculator traversed each lemmatized document, summing the corresponding polarity values of matched terms. This process was repeated for both corporate reports and media texts. The sentiment score for each company was then normalized by dividing the total polarity-weighted sum by the *Total Number of Tokens*, yielding a comparative sentiment score:

$$\text{Comparative Score} = \frac{\text{Total Score}}{\text{Total number of Tokens}} \quad (1)$$

This measure captures the relative tone of sustainability language and enables direct comparison between a company's self-presentation and its portrayal in the media.

2.3. Consistency Assessment

In the continuation of the work, the correlation of Pearson at the level of country, industry and company size was calculated, in order to assess the consistency between the comparative scores of the resulting companies based on the sustainability reports and on the basis of those reflected in the news. This analysis was conducted to understand the extent to which companies' official statements regarding their sustainability practices align with public perception. The Pearson coefficient [38] is one of the most widely used methods for evaluating the relationship between variables, due to its simplicity and interpretability, and is based on the following formula:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (2)$$

where X_i and Y_i represent the comparative sentiment scores derived from company reports and media coverage, respectively. High positive correlations indicate strong alignment between official and public narratives, suggesting credible communication, whereas low or negative values imply potential reputational discrepancies or greenwashing tendencies.

2.4. Topic Modeling and Term Weighting

To gain an even deeper insight into the thematic content of sustainability reports, Latent Dirichlet Allocation (LDA) was used, a topic modeling technique that helps discover latent topics or themes within the reports [39].

LDA is a probabilistic generative model that assumes that each document is a combination of several topics, and each topic is a combination of terms. The formula on which LDA is based is:

$$P(w_d) = \prod_{n=1}^N \sum_{k=1}^K P(w_{d,n} | z_{d,n} = k, \beta) P(z_{d,n} = k | \theta_d) \quad (3)$$

where K denotes the number of latent topics and $P(w_{d,n} | z_{d,n} = k, \beta)$ represents the probability of observing word n in document d given that it is assigned to topic k and the topic distribution β . Furthermore, $P(z_{d,n} = k | \theta_d)$ denotes the probability that topic k is selected for word n in document d , given the document-specific topic distribution θ_d .

Ref. [40] devised a statistical interpretation of term specificity called Inverse Document Frequency (IDF), which has become an essential pillar of term weighting. This idea has evolved over time into TF-IDF, a statistical measure that assesses how relevant a word is to a document in a collection of documents.

TF-IDF has many possibilities of use, the most important being in automatic text analysis, being a very useful method for notating words in machine learning algorithms for natural language processing (NLP).

TF-IDF vectorization transforms preprocessed textual data into a numerical feature matrix and adjusts the frequency of words so that common words in all documents are diminished and specific words are amplified. All ESG focus scores were normalized using Min–Max scaling, rescaling values to the interval [0, 0.99]. This transformation preserves relative differences in term emphasis across firms while ensuring comparability across documents and facilitating aggregation into the composite Greenwashing Severity Index. The TF-IDF methodology combines the two elements TF and IDF , as follows:

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D)$$

where $TF(t, d)$ is:

$$TF(t, d) = \frac{f_{t,d}}{N_d},$$

The variable represents the frequency of the term t in document d , where $f_{t,d}$ is the number of occurrences of term t in document d , and N_d is the total number of terms in document d .

Furthermore, $IDF(t, D)$ is defined as:

$$IDF(t, D) = \log \left(\frac{N_D}{n_t} \right)$$

where N_D denotes the total number of documents in the collection D , and n_t is the number of documents containing the term t .

2.5. ESG Focus Scores

Next, a methodology based on ESG focus scores was used, which combines the raw frequencies of the terms and the TF-IDF technique to provide an accurate measure of the importance of the terms in the sustainability reports. A function has been created that calculates the focus score for each ESG (environmental, social and governance) category. It works by counting the occurrences of relevant keywords in the text, using the three ESG dictionaries created beforehand, and dividing this number by the total number of terms in the text. The formula used is:

$$\text{Focus Score} = \frac{\text{Total Number of Keyword Occurrences}}{\text{Total Number of Tokens}}$$

Raw scores have been normalized and scaled to ensure comparability between documents. Normalization adjusts the values so that they are between 0 and 0.99.

Also, to get an overall assessment of each company's ESG commitment, the average focus score was calculated by aggregating the normalized scores for each category, as follows:

$$\text{Average Score} = \frac{E_{focus} + S_{focus} + G_{focus}}{3}$$

For each document, the environmental, social, and governance focus scores were then calculated by aggregating the $TF-IDF$ scores of the keywords in each set, as follows:

$$E_{tfidf} = \sum_{t \in E} TF - IDF(t, d, D) \quad (\text{Environmental component})$$

$$S_{tfidf} = \sum_{t \in S} TF - IDF(t, d, D) \quad (\text{Social component})$$

$$G_{tfidf} = \sum_{t \in G} TF - IDF(t, d, D) \quad (\text{Governance component})$$

The E_{tfidf} , S_{tfidf} , and G_{tfidf} scores provide an accurate and contextualized measure of the importance of ESG terms in each document. These scores reflect how frequently and how relevant specific environmental, social and governance terms appear in that document, relative to the entire set of documents.

2.6. Construction of the Greenwashing Severity Index (GSI)

The core of this research lies in the construction of the Greenwashing Severity Index (GSI), a composite indicator designed to quantify discrepancies between reported sustainability commitments and the tone of actual communication.

The calculation involves an iterative process through predefined sustainability indicators. For each indicator, *TF-IDF* scores were used to assess its importance in a report. By aggregating these scores for all indicators, the GSI was derived, which has the following formula:

$$GSI = (\omega_E \cdot E_{tfidf}) + (\omega_S \cdot S_{tfidf}) + (\omega_G \cdot G_{tfidf})$$

where ω_E , ω_S , ω_G are the weights assigned to each ESG focus score and are:

$$\omega_E = \omega_S = \omega_G = \frac{1}{3}$$

Assigning equal weights reflects an unbiased approach, ensuring that each ESG (Environmental, Social, Governance) dimension is considered equally important in the overall assessment. This is crucial to avoid underestimating or overestimating a certain dimension to the detriment of others. Higher index values indicate a greater likelihood of greenwashing or misalignment between discourse and practice, whereas lower values signify consistency and integrity in sustainability communication.

To assess the robustness of the GSI, sensitivity analyses were conducted using alternative weighting schemes and partial index constructions. Specifically, the index was recalculated using unequal weights that emphasized one ESG dimension at a time, as well as by excluding individual dimensions from the aggregation. Across these specifications, firm rankings and aggregate patterns by country, industry, and firm size remained qualitatively stable. This indicates that the GSI is not driven by any single ESG pillar but captures a systemic discrepancy in sustainability communication.

3. Results

As outlined in the previous section, the initial stage of the analysis followed data collection and preprocessing and involved examining the correlations between the comparative sentiment scores derived from companies' sustainability reports and those obtained from news coverage related to the same firms. The main objective was to assess the degree of alignment between corporate self-representation in official sustainability disclosures and public perception as reflected in external media narratives. Correlations were calculated by grouping firms according to their country of origin, industry sector, and company size. The results of the country-level correlations are presented in Table 1.

Table 1. Country correlation results.

Country	Correlation
Bulgaria	0.0797
Czech Republic	0.3804
Croatia	0.0008
Germany	0.1494
Lithuania	0.0621
Poland	0.1787

Romania	0.2129
Serbia	0.1118
Slovakia	0.1053
Hungary	0.0076

Note: The table shows the correlations between sustainability reports and news, grouped by country. These were calculated using the Pearson correlation coefficient. Values close to 1 indicate a strong correlation, suggesting alignment between the two variables, while values close to 0 indicate a lack of correlation, i.e., dissonance.

The highest level of consistency is observed in the case of the Czech Republic, with a correlation of 0.38, indicating a moderate and statistically significant alignment between officially communicated information and public perceptions. While modest in magnitude, this correlation is meaningful in a cross-firm, text-based setting characterized by heterogeneous reporting practices. The result is robust to alternative correlation measures, yielding consistent qualitative patterns. This result implies a higher degree of transparency and a lower likelihood of greenwashing. Romania and Poland display moderate correlation coefficients of 0.21 and 0.17, respectively, reflecting partial alignment and a moderate risk of greenwashing. In contrast, Croatia and Hungary record near-zero correlations (0.0008 and 0.0076), signaling a pronounced discrepancy between sustainability reports and media coverage, which points to a higher probability of greenwashing practices.

Table 2 contains the results of the correlations between the two comparative scores analyzed by industry type. The lowest correlation coefficients are in the case of the finance industry, equal to 0.013 and the alcoholic and non-alcoholic beverages industry, equal to 0.018, highlighting the fact that these two industries, in the case of Central and Eastern Europe, are more susceptible to greenwashing practices.

Table 2. Results of industry correlations.

Industry	Correlation
Agriculture	0.1232
Aviation	0.6856
Household appliances	0.1724
Cars	0.2613
Alcoholic and non-alcoholic beverages	0.0188
Construction	0.3098
Online commerce	0.3728
Energy	0.1059
Engineering & Technology	0.2681
Finance	0.0136
The Preparation of Food	0.2419
Health	0.8901
EN	0.9165
Insurance	0.3861
Manufacturing	0.1365
Oil & Gas	0.3739
Pharmaceutical	0.6615
Public transport	0.9145
Tobacco	0.3004
Sales	0.2038

Note: The table shows the correlations between sustainability reports and news, grouped by industry. These were calculated using the Pearson correlation coefficient. Values close to 1 indicate a

strong correlation, suggesting alignment between the two variables, while values close to 0 indicate a lack of correlation, i.e., dissonance.

The industries with the highest correlation coefficients are: IT (0.916), Health (0.89) and Public Transport (0.914), which shows that the information communicated in the sustainability reports by companies in these industries reflects reality, thus reducing the risk of almost total greenwashing.

As shown in Table 3, large companies record the highest correlation coefficient (0.89), indicating a substantially lower risk of greenwashing. This outcome can be attributed to the greater visibility, public scrutiny, and regulatory oversight that typically accompany large-scale operations, which enhance transparency and the verifiability of disclosed information. Beyond visibility effects, firm size is also closely associated with the strength of internal governance structures. Large firms generally benefit from more formalized ESG governance arrangements, including dedicated sustainability teams, internal audit and compliance functions, and standardized reporting processes aligned with international frameworks. These internal controls reduce information asymmetries and increase consistency between sustainability disclosures and actual corporate practices.

In contrast, medium-sized and small companies exhibit considerably lower correlations, 0.019 and 0.050, respectively, reflecting a higher degree of misalignment between reported sustainability claims and external perceptions, and suggesting a greater susceptibility to greenwashing practices. This pattern may be explained by more limited governance capacity and resource constraints, which often restrict the ability of smaller firms to implement robust monitoring, verification, and reporting mechanisms. Sustainability communication in such firms may therefore rely more on symbolic or selective disclosure strategies rather than on verifiable performance-based reporting.

Additionally, lower levels of regulatory supervision and media attention reduce external accountability pressures for medium-sized and small firms. The weaker threat of reputational or regulatory sanctions lowers the expected cost of inconsistent sustainability claims, allowing dissonance between corporate narratives and external assessments to persist. Overall, these findings highlight firm size as a critical determinant of greenwashing risk, operating through both internal governance quality and the intensity of external scrutiny.

Table 3. Results of company-size correlations.

Size	Correlation
Large Company	0.8901
Medium company	0.0191
Small Company	0.0508

Note: The table shows the correlations between sustainability reports and news, grouped by company size. These were calculated using the Pearson correlation coefficient. Values close to 1 indicate a strong correlation, suggesting alignment between the two variables, while values close to 0 indicate a lack of correlation, i.e., dissonance.

Three word clouds were generated for the Environmental, Social, and Governance (ESG) dimensions to show the most frequently used terms in the sustainability reports. These visualizations highlight which concepts appear most often in company disclosures and indicate the areas that receive the most attention in corporate reporting.

Figure 1 shows the word cloud analysis for environmental keywords, using the E dictionary. The most prominent terms, such as “sustainability report”, “energy”, “sustainable development”, “climate change”, “environmental impact” or “renewable energy”, indicate the strong emphasis placed by companies in their reporting on environmental responsibility.

The charts shown in Figure 4 resulting from the application of the LDA model show the distribution of keyword frequency for each topic identified in the sustainability reports. Each graph illustrates the most relevant words associated with a particular topic, highlighting the predominant themes of the analyzed text.

Analyzing the following figure, it can be clearly seen that Topic 0 is the topic that focuses on the environment, the most frequent words concretely highlighting sustainability issues (“sustainability”, “environment”, “climate”). This pattern suggests that environmental protection remains one of the primary priorities in corporate communication.

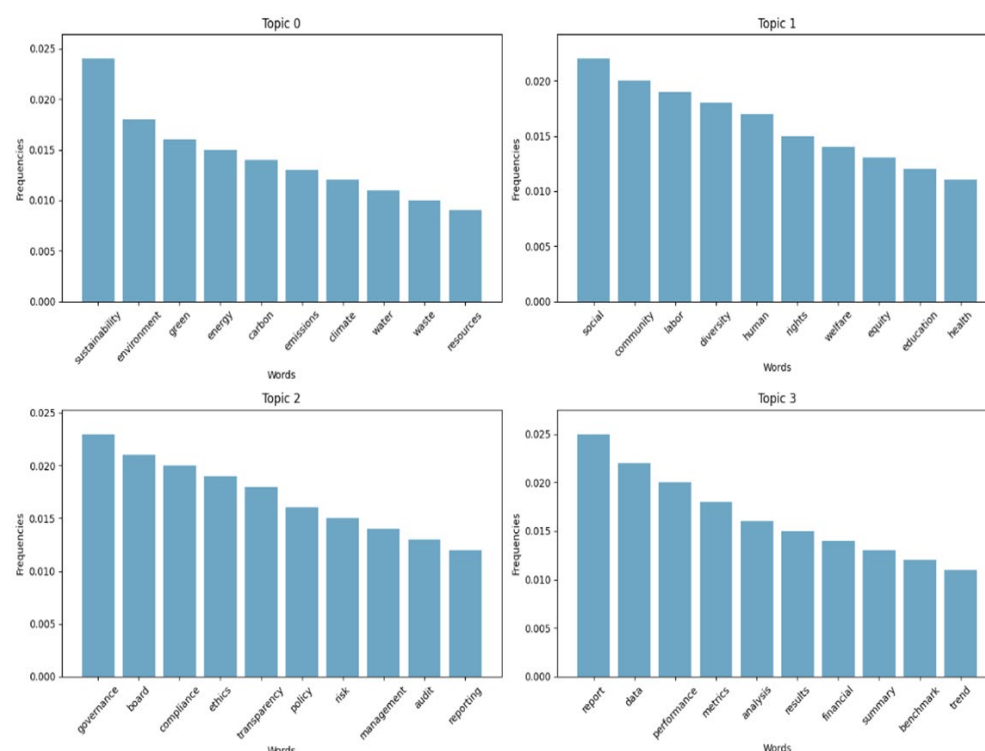


Figure 4. LDA Topic Modeling.

Topic 1 highlights the social dimension, reflected in words such as “social,” “community,” “human,” and “health,” pointing to an emphasis on social responsibility and stakeholder well-being. Topic 2 focuses on governance, as shown by terms including “management,” “risk,” and “governance,” underlining the role of oversight, ethical leadership, and internal control.

Topic 3 captures the idea of performance measurement and reporting, through words such as “analysis” and “performance,” demonstrating companies’ interest in evaluating and communicating their progress.

The identification of these topics, with the help of the LDA analysis on companies’ sustainability reports, reflects the priorities of the organizations included in the analysis. Therefore, we can name the identified topics as follows: Topic 0 is “The environment”; Topic 1 is “Social component”; Topic 2 stands for “Company Management” and Topic 3 is “Performance Reporting.”

As expected, the major focus of companies is on the environment, social responsibility and corporate governance, which confirms the compliance with ESG criteria by the companies analyzed.

It can also be noticed that organizations also pay special attention to monitoring, evaluating and communicating their performance. This may indicate that the steps taken for the purpose of final disclosure are included in the sustainability reports.

In the next part of the research, the analysis of the greenwashing severity index is presented, calculated based on the focus scores E, S and G. After determining the index for each company of the 204 included in the analysis, descriptive statistics were d based on it.

Figure 5 illustrates the distribution of GSI values in the dataset. The horizontal axis represents the index values, and the vertical axis represents the frequency of occurrence. Moreover, to formally assess the distributional properties of the GSI, skewness, kurtosis, and normality tests were conducted. The GSI exhibits pronounced negative skewness (skewness = -1.66), indicating a long-left tail with relatively few firms displaying very low greenwashing severity and a strong concentration of observations at moderate-to-high index values. Excess kurtosis is substantial (6.56), revealing a leptokurtic distribution with heavy tails and marked clustering around the central range. A Shapiro–Wilk test strongly rejects the null hypothesis of normality ($W = 0.889$, $p < 0.001$), confirming that the GSI distribution deviates significantly from Gaussian assumptions. These results indicate that greenwashing severity is asymmetrically distributed across firms, with low-severity cases being comparatively rare.

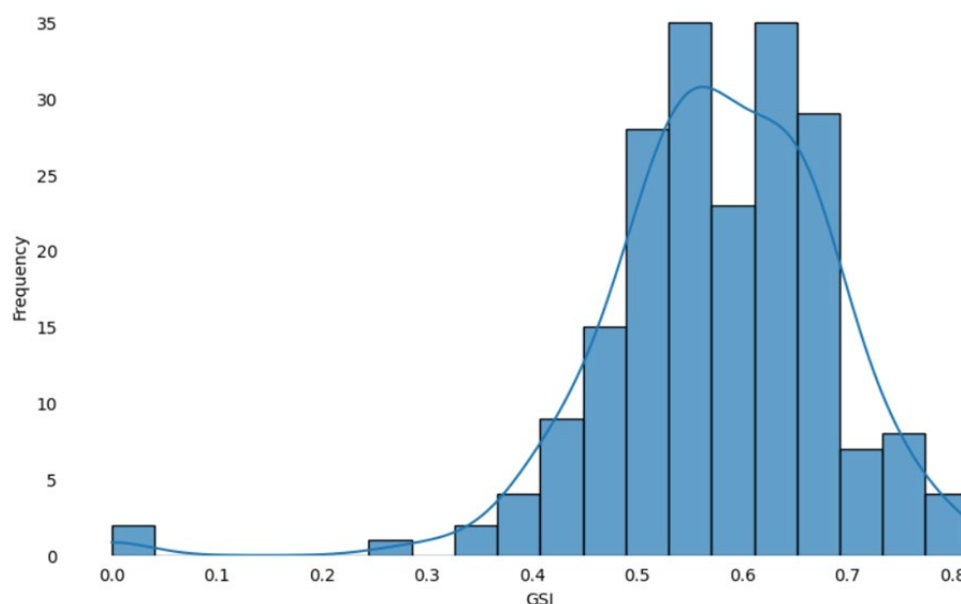


Figure 5. Distribution of the Greenwashing Severity Index for the companies analyzed.

It can be noted that the highest frequency is in the range of 0.5 and 0.7; there are very few companies with very low GSI values (close to 0) and a moderate number of companies with very high values (close to 0.8), which proves that extreme greenwashing practices (either very low or very high) are less common.

Table 5 presents descriptive statistics that summarize the central tendencies and dispersion of the analyzed variables. The mean values reflect the general magnitude of each indicator, providing an overview of the overall tendencies within the dataset. The average Greenwashing Severity Index (GSI) of approximately 0.57 indicates a moderate level of greenwashing among the analyzed firms, consistent with the distribution pattern observed in the histogram. This suggests that, across the sample, corporate sustainability communication often includes elements of overstatement or selective disclosure rather than outright fabrication.

The standard deviation values reflect the degree of variability across the analyzed indicators, indicating how dispersed the observations are around their respective means. For instance, the standard deviation of approximately 0.12 for the environmental component (E Washing) denotes a moderate spread in environmental focus scores, suggesting some heterogeneity in the extent of environmental emphasis among firms. The percentile

values, particularly the 25th and 75th percentiles (Q1 and Q3), offer further insight into the data distribution. The 25th percentile of 0.55 for the governance component (G Washing) indicates that one-quarter of the companies exhibit governance-related scores below this threshold, highlighting uneven consistency in governance-related communication practices across the sample.

Table 5. Descriptive statistics of the analyzed indicators.

Indicators	Min	Q1	Median	Intercede	Q3	Max	Standard Deviation
GSI	0.00063	0.5200	0.5745	0.5765	0.6510	0.8150	0.1111
E Washing	0.00067	0.3667	0.4540	0.4474	0.5163	0.8397	0.1232
S Washing	0.00042	0.6026	0.6746	0.6566	0.7370	0.8654	0.1207
G Washing	0.00029	0.5536	0.6275	0.6254	0.7090	0.9003	0.1267

Note: This table presents descriptive statistics for key numeric variables in the dataset (e.g., mean standard deviation, minimum, maximum, and distribution percentiles). It shows the central trends, variability and distribution of data on the severity index (GSI), the disinformation factor for the environment (E Washing), for the social factor (S Washing) and for the governance factor (G Washing).

Table 6 presents the average severity index grouped by country of residence for the analyzed companies. Given that all GSI values exceed 0.5, greenwashing appears to be a widespread issue across all examined countries, indicating that none can be regarded as fully transparent or entirely accurate in their sustainability reporting. Although every country displays a moderate to significant degree of greenwashing, the intensity varies across cases. Germany and Romania register the highest GSI values, suggesting a more pronounced tendency toward overstated sustainability communication, while Lithuania records the lowest value, reflecting relatively greater consistency between discourse and practice. Germany's index of 0.6086 represents the highest level among the countries studied, implying a stronger potential for misleading environmental claims in corporate disclosures. Across the ESG dimensions, most countries place greater emphasis on social and governance themes than on environmental concerns, indicating that corporate sustainability narratives in the region tend to prioritize organizational ethics and social engagement over ecological accountability.

Table 6. GSI results grouped by country.

Country	GSI	E Washing	S Washing	G Washing
Germany	0.6086	0.5061	0.6602	0.6595
Romania	0.5970	0.4560	0.6868	0.6482
Hungary	0.5943	0.4843	0.6579	0.6407
Serbia	0.5908	0.4748	0.6627	0.6350
Slovakia	0.5747	0.4408	0.6656	0.6177
Czech Republic	0.5723	0.4323	0.6601	0.6244
Bulgaria	0.5654	0.4227	0.6500	0.6235
Croatia	0.5621	0.4207	0.6494	0.6111
Poland	0.5414	0.4203	0.6185	0.5853
Lithuania	0.5157	0.3880	0.5953	0.5636

Note: This table shows the GSI for different countries. GSI measures the potential of green disinformation practices within companies. The table also includes the corresponding environmental (E), social (S) and governance (G) disinformation scores.

Overall, results indicate moderate but widespread greenwashing across all analyzed countries, with an average Greenwashing Severity Index (GSI) of 0.57. This suggests that while most firms in Central and Eastern Europe publicly endorse sustainability, their

disclosures often involve selective reporting and rhetorical exaggeration rather than out-right fabrication.

The Czech Republic displays the highest consistency between sustainability reports and media narratives, while Germany and Romania record the strongest misalignment, signaling a higher risk of deceptive communication.

The analysis of GSI scores by industry as shown in Table 7 reveals that greenwashing is a widespread problem in all sectors analyzed. The e-commerce, finance, and aviation industries have the highest GSI scores (0.64, 0.63, and 0.61, respectively), suggesting that these industries are most likely to distort the truth in sustainability reporting. On the other hand, the construction industry, although it has the lowest GSI score (0.49), is not completely devoid of the possibility of greenwashing practices.

Table 7. GSI results grouped by industry.

Industry	GSI	E Washing	S Washing	G Washing
Agriculture	0.5709	0.4179	0.6571	0.6378
Aviation	0.6148	0.5323	0.6571	0.6550
Household appliances	0.5599	0.3777	0.6533	0.6486
Cars	0.5492	0.4172	0.6289	0.6014
Alcoholic and non-alcoholic beverages	0.5172	0.4283	0.5543	0.5690
Construction	0.4953	0.4597	0.5416	0.4847
Online commerce	0.6496	0.4826	0.7488	0.7174
Energy	0.5486	0.4077	0.6450	0.5931
Engineering & Technology	0.5850	0.4869	0.6527	0.6155
Finance	0.6302	0.5142	0.6978	0.6786
The Preparation of Food	0.5192	0.4206	0.5744	0.5627
Health	0.5905	0.5160	0.6406	0.6149
EN	0.5917	0.4709	0.6727	0.6316
Insurance	0.6052	0.4938	0.6786	0.6432
Manufacturing	0.5481	0.4191	0.6275	0.5978
Oil & Gas	0.5777	0.4427	0.6692	0.6212
Pharmaceutical	0.5439	0.4156	0.6395	0.5766
Public transport	0.5542	0.3698	0.6743	0.6185
Tobacco	0.5912	0.4348	0.6990	0.6400
Sales	0.5058	0.4041	0.5604	0.5529

Note: This table shows the GSI for the industries analyzed. GSI measures the potential of green disinformation practices within companies. The table also includes the corresponding environmental (E), social (S) and governance (G) disinformation scores.

Overall, as expected from the analysis of the country-grouped index, the focus is on social and governance aspects, reinforcing the idea of a possible attempt to distract attention from environmental issues.

Industry-level analysis reveals that online commerce, finance, and aviation exhibit the highest levels of greenwashing, reflecting stronger tendencies toward reputational signaling and selective disclosure. In contrast, IT, healthcare, and public transport demonstrate greater alignment between internal reporting and external perception, suggesting higher transparency and credibility.

From the information presented in the previous table, it can be deduced that the probability of communicating false information related to the sustainable practices of companies is higher in the case of medium and small companies, than in the case of large ones. This can be closely linked to the higher level of control and supervision, as well as

the greater resources for the implementation of authentic sustainable practices that large companies have at their disposal.

On the other hand, as shown in Table 8, small companies have the highest values, both of the GSI (0.59) and of the E, S, and G indices (0.46, 0.66, and 0.64, respectively), indicating a strong tendency to exaggerate their discourse in sustainability reports.

Table 8. GSI results grouped by company size.

Size	GSI	E Washing	S Washing	G Washing
Large Company	0.4848	0.2588	0.6472	0.5484
Medium Company	0.5584	0.4348	0.6360	0.6046
Small Company	0.5928	0.4677	0.6644	0.6463

Note: This table shows the GSI for company sizes. GSI measures the potential of green disinformation practices within companies. The table also includes the corresponding environmental (E), social (S) and governance (G) disinformation scores.

Firm size emerges as a key determinant of transparency. Large companies show stronger alignment ($r = 0.89$) and lower GSI values (0.49), consistent with higher visibility and regulatory scrutiny. Small and medium enterprises, however, record significantly higher greenwashing scores, underscoring capacity and oversight gaps in ESG reporting.

The correlations (Table 9) between the different variables analyzed in the paper reveal important relationships between the ESG focus and greenwashing trends. The overall Greenwashing Severity Index (GSI) is strongly correlated with environmental (E Washing, 0.82), social (S Washing, 0.91) and governance (G Washing, 0.96) greenwashing, indicating that firms that tend to overdo it in one ESG dimension are inclined to do the same in the rest.

Table 9. Correlation matrix.

	GSI	E Focus	S Focus	G Focus	E Washing	S Washing	G Washing	ESG Focus
GSI	1.00							
E Focus	0.51	1.00						
S Focus	0.46	0.26	1.00					
G Focus	0.57	0.37	0.88	1.00				
E Washing	0.82	0.76	0.09	0.26	1.00			
S Washing	0.91	0.28	0.65	0.63	0.54	1.00		
G Washing	0.96	0.34	0.51	0.64	0.66	0.92	1.00	
ESG Focus	0.63	0.65	0.88	0.92	0.44	0.64	0.61	1.00

Note: This table shows the correlation coefficients between the indicators used, such as ESG focus scores, Greenwashing index (GSI), and disinformation scores for environment (E), social (S), and governance (G). Correlations provide insight into potential relationships between these values.

Environmental Focus (E Focus) has a moderate correlation with the greenwashing index (0.51) and is strongly correlated with environmental greenwashing (0.76), suggesting that the accentuation of environmental aspects can often be accompanied by the phenomenon of false propaganda.

The social focus (S Focus) has strong correlations with social greenwashing (0.65) and governance greenwashing (G Focus, 0.88), and the focus on governance is closely related to governance greenwashing (G Washing, 0.64).

Similarly, the average overall ESG focus score per company (ESG Focus) has quite strong correlations with each dimension of focus and greenwashing, indicating that an overall emphasis on ESG in sustainability reporting is associated with an increase in false

propaganda trends across all dimensions. The table demonstrates that while companies are stepping up their ESG reporting efforts, they tend to overdo it and practice greenwashing.

Strong correlations among environmental, social, and governance greenwashing dimensions suggest that firms tend to exaggerate across multiple ESG areas simultaneously. This confirms that greenwashing is not isolated to environmental claims but a systemic communication strategy spanning all sustainability dimensions.

Figure 6 plots Refinitiv's ESG [41] composite and the three pillar scores for each country in our sample. Scores span a wide range on the 0–100 scale, and the ranking is not stable across pillars. Countries that sit near the top of the composite such as Germany and the Czech Republic do not consistently lead on all three sub-dimensions, while countries at the lower end of the composite—such as Serbia and Bulgaria—occasionally post comparatively stronger results on a given pillar. These shifts in relative position from the composite to Environment, Social, and Governance underscore that national sustainability regimes are genuinely multidimensional: the institutional conditions firms face depend on which pillar is considered, and no single pillar fully summarizes the country context.

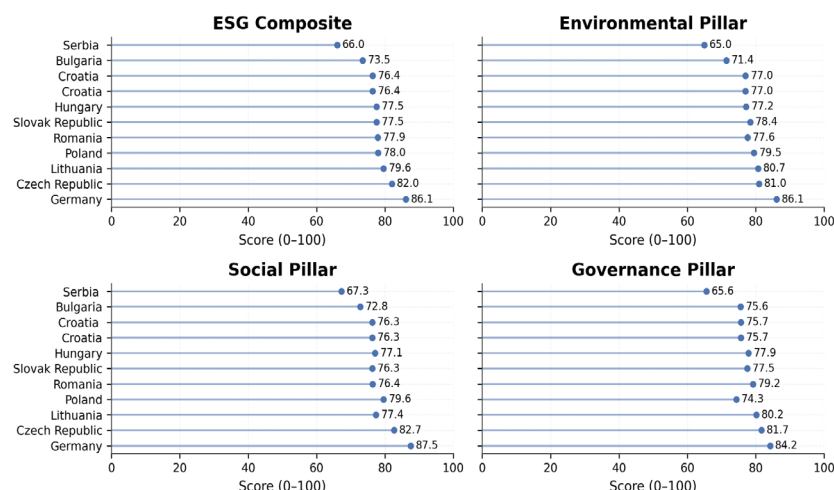


Figure 6. Refinitiv's ESG composite and E/S/G pillar scores by country.

Figure 7 relates these country scores to the country-level mean of the Greenwashing Severity Index (GSI). In the environmental panel the fitted line slopes slightly downward, which suggests that stronger environmental conditions at the national level are associated with a lower average GSI. Lithuania is illustrative of this pattern, combining a comparatively higher environmental score with a lower mean GSI. The association is, however, shallow and sensitive to influential observations; Germany, for example, exhibits both a high environmental score and a relatively high country mean of GSI, reminding us that within-country firm heterogeneity remains large.

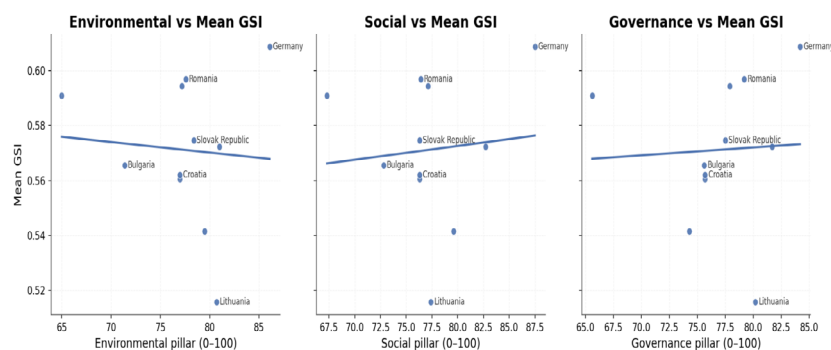


Figure 7. Country scores to the country-level mean GSI.

The social panel shows the opposite tendency. The fitted line is mildly upward sloping, indicating a weak positive relationship between national social scores and the country mean of GSI. One interpretation is that stronger social frameworks coincide with more extensive ESG communication in corporate narratives, which our text-based index may flag when the intensity of disclosure outpaces alignment with external perceptions. Because the slope is modest and the points are few, this pattern should be read as descriptive rather than conclusive.

A similar picture emerges for governance. The governance panel does not reveal a negative association between national governance quality and mean GSI; if anything, the fitted line is slightly positive. Germany again stands out with both high governance scores and higher mean GSI, underscoring that a favorable governance setting at the national level does not automatically translate into uniformly lower greenwashing severity among firms. Together, the three panels emphasize heterogeneity both across countries and within them, and they caution against attributing firm-level communication behavior solely to country-level institutions.

Taken together, it is suggested that, within this CEE sample, country context matters less than firm-specific and sectoral factors for explaining greenwashing severity. National ESG frameworks vary meaningfully, but their translation into lower firm-level GSI is limited on average. Any tentative negative association appears confined to the environmental pillar and is small relative to the overall dispersion in GSI. These findings motivate the subsequent discussion to focus on mechanisms operating at the firm and industry levels such as disclosure strategies, media exposure, and managerial discretion while acknowledging that the country environment sets the broader backdrop against which those choices are made.

To formally assess whether higher disclosure asymmetry is associated with more severe greenwashing, a propensity score-based matching strategy was employed. The treatment was defined as high ESG focus imbalance, corresponding to firms with above-median dispersion across their normalized environmental, social, and governance focus indicators. Propensity scores were estimated using a logistic regression model including country-level ESG pillar scores, firm size, as well as country and industry fixed effects, thereby accounting for both institutional and firm-level confounders.

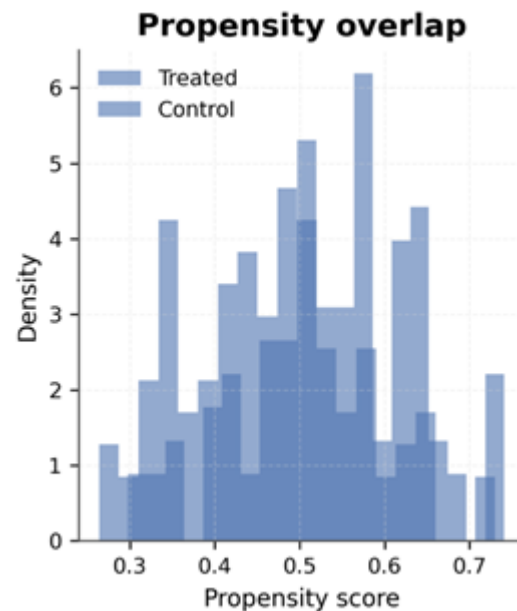
As a baseline specification, nearest neighbor matching on the propensity score was implemented without replacement, applying a caliper of 0.09 on the logit of the propensity score in line with standard recommendations to reduce poor matches [42]. Covariate balance was assessed using standardized mean differences (SMDs) [43], indicating improved alignment for firm size and country-level ESG scores following matching, although some residual imbalance remained for governance-related covariates.

To evaluate the robustness of the findings, the analysis was extended using Mahalanobis distance matching within the same propensity score caliper. This alternative specification minimizes multivariate distance across standardized covariates while maintaining common support defined by the propensity score, resulting in a larger matched sample and stricter comparability across firms.

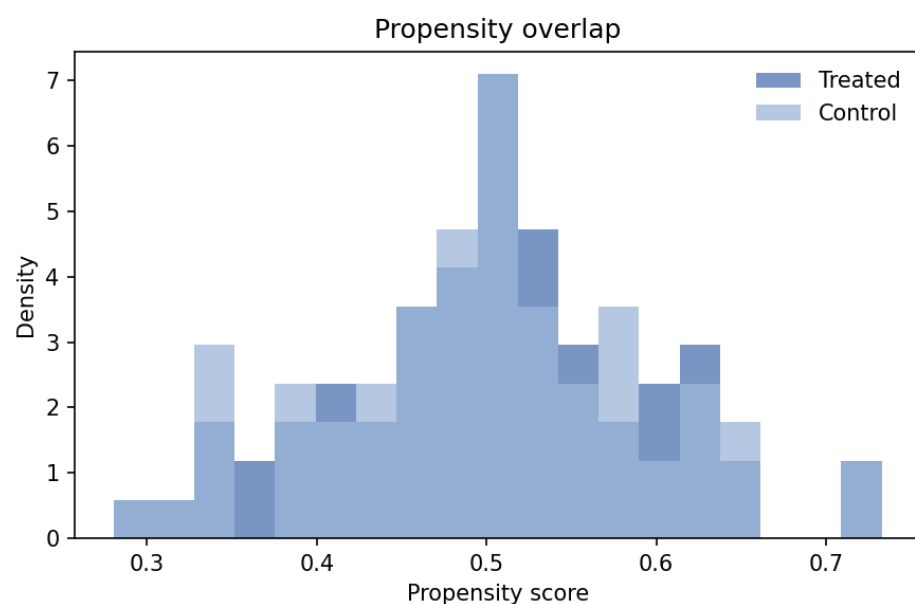
Results from both matching strategies are presented in Tables 10 and 11, with corresponding propensity score distributions for treated and control firms shown in Figures 8 and 9. Under nearest neighbor propensity score matching, the estimated Average Treatment Effect on the Treated (ATT) is 0.048 ($p = 0.077$; 48 matched pairs). The Mahalanobis-based specification yields a comparable estimate of 0.040 ($p = 0.064$; 71 matched pairs). Although both estimates fall marginally below conventional significance thresholds, their consistent positive magnitude and direction across matching techniques indicate a systematic association between ESG focus imbalance and greenwashing severity.

Table 10. Propensity Score Matching (Nearest Neighbor) estimates for the effect of high focus imbalance on the Greenwashing Severity Index (GSI).

Variable	ATT	95% ci (Lower, Upper)	<i>p</i> -Value	Matched Pairs	Caliper (Logit)
High focus imbalance (treatment effect on GSI)	0.048	(−0.004, 0.100)	0.077	48	0.090

**Figure 8.** Propensity score distribution for treated and control firms (Nearest Neighbor).**Table 11.** Propensity Score Matching (Mahalanobis) estimates for the effect of high focus imbalance on the Greenwashing Severity Index (GSI).

Variable	ATT	95% ci (Lower, Upper)	<i>p</i> -Value	Matched Pairs	Caliper (Logit)
High focus imbalance (treatment effect on GSI)	0.0403	(−0.0017, 0.0823)	0.0640	71	0.090

**Figure 9.** Propensity score distribution for treated and control firms (Mahalanobis).

Overall, the results suggest that selective emphasis on individual ESG dimensions is associated with higher Greenwashing Severity Index (GSI) values even after controlling for observable firm- and country-level characteristics. The robustness of the findings across alternative matching procedures supports the interpretation that disclosure asymmetry reflects strategic reporting behavior rather than purely structural or institutional conditions. These findings reinforce the need for more integrated ESG reporting frameworks and enhanced regulatory oversight aimed at reducing fragmented disclosure practices and strengthening the credibility of sustainability communication.

To evaluate the robustness of the GSI, we conducted a set of sensitivity analyses examining alternative weighting schemes and partial index constructions. Table 12 reports Spearman rank correlations between the baseline equal-weighted GSI and alternative specifications.

When higher weights were assigned to individual ESG dimensions (environmental-, social-, or governance-dominant indices), rank correlations with the baseline GSI remained high, ranging between 0.94 and 0.97. This indicates that firm rankings and relative greenwashing severity are not driven by any single ESG pillar.

Additional robustness checks were performed by constructing partial indices that exclude one ESG dimension at a time. Although absolute index values changed mechanically, rank correlations with the full GSI remained above 0.88 in all cases. Importantly, the qualitative findings were preserved: industries identified as high-risk (finance, online commerce, aviation) continued to exhibit elevated greenwashing severity, while large firms consistently showed lower index values than small and medium-sized enterprises.

Overall, these results confirm that the GSI provides a stable and balanced summary measure of greenwashing severity rather than being sensitive to arbitrary weighting choices or dimensional dominance.

Table 12. Robustness of the GSI under alternative specifications.

Specification	Description	Spearman Rank Correlation with Baseline GSI
Baseline GSI	Equal weights ($E = S = G = 1/3$)	1.000
Environmental-dominant	$E = 0.50, S = 0.25, G = 0.25$	0.842
Social-dominant	$E = 0.25, S = 0.50, G = 0.25$	0.941
Governance-dominant	$E = 0.25, S = 0.25, G = 0.50$	0.966
GSI (−E)	Environmental dimension excluded	0.788
GSI (−S)	Social dimension excluded	0.908
GSI (−G)	Governance dimension excluded	0.922

Note: Spearman correlations assess the stability of firm rankings relative to the baseline equal-weighted GSI.

4. Discussion

This study contributes to the literature on sustainability governance by demonstrating how digitalization, understood as the systematic use of data-driven and computational tools for information processing and oversight, can enhance transparency and accountability in corporate ESG reporting.

Rather than treating greenwashing as a phenomenon confined to environmental claims or marketing practices, the results support a broader interpretation of greenwashing as a multidimensional governance issue spanning environmental, social, and governance dimensions. This interpretation aligns with recent conceptual work emphasizing the gap between reported and substantive ESG performance [19,29]. By operationalizing greenwashing through an ESG-integrated, NLP-based Greenwashing Severity Index (GSI), the study

reframes greenwashing as a relational outcome shaped by institutional capacity, monitoring intensity, and stakeholder scrutiny rather than by disclosure volume alone.

Focusing on Central and Eastern Europe, the analysis reveals a heterogeneous landscape of ESG transparency. The Czech Republic exhibits the strongest alignment between sustainability reports and public discourse, consistent with institutional perspectives highlighting the role of regulatory quality and enforcement capacity in shaping disclosure integrity [17,22]. By contrast, higher GSI values observed in Germany and Romania indicate more pronounced discrepancies between corporate self-representation and external perceptions. Importantly, this pattern does not imply weaker regulation per se. Instead, it suggests that regulatory maturity may shape the form of disclosure more than its substance, allowing firms in highly regulated environments to comply formally while engaging in increasingly sophisticated narrative optimization. This finding extends legitimacy theory by illustrating how symbolic compliance may persist even under advanced regulatory regimes when interpretive oversight remains limited.

At the industry level, the results further support a governance-based interpretation of greenwashing. Sectors characterized by high reputational exposure and direct stakeholder interaction such as IT, healthcare, and public transport display strong coherence between internal sustainability narratives and external media discourse. While this may reflect genuinely stronger accountability pressures, it may also partly stem from sector-specific reporting conventions and media homogeneity, where standardized sustainability language dominates both corporate and journalistic narratives. These results caution against interpreting high textual alignment as unequivocal evidence of superior transparency.

Conversely, sectors such as online commerce, finance, and aviation exhibit higher GSI values, indicating a greater propensity for strategic sustainability rhetoric. This finding is consistent with prior evidence that market-driven sustainability communication, when weakly monitored, tends to evolve toward symbolic signaling rather than substantive performance disclosure [4,9]. The results suggest that greenwashing risk is shaped not only by environmental impact but also by reputational incentives and monitoring asymmetries across industries.

Firm size emerges as a critical moderating factor. Large firms show stronger alignment between internal sustainability narratives and external assessments, reflecting both more developed internal governance structures and higher levels of regulatory and media scrutiny. This finding is consistent with earlier studies linking governance quality and ethical oversight to lower levels of deceptive ESG reporting [16,20]. Smaller firms, by contrast, exhibit higher GSI values and weaker narrative alignment, likely reflecting fewer disclosure obligations, limited institutional resources, and lower external scrutiny. In this sense, greenwashing among smaller firms appears less as deliberate deception and more as a by-product of structural constraints, reinforcing observations of “formalistic compliance” in the CEE region [22].

From a methodological perspective, the consistency between sentiment correlations, ESG focus indicators, and GSI values supports the robustness of the proposed NLP-based framework. By combining lexicon-based sentiment analysis, TF-IDF weighting, and topic modeling, the GSI captures greenwashing as a relational and multidimensional phenomenon rather than a binary outcome, extending prior NLP-based approaches to ESG assessment [15,30].

The propensity score matching analysis provides further insight into the mechanisms underlying greenwashing. Firms exhibiting imbalanced emphasis across ESG dimensions display systematically higher GSI values than otherwise comparable firms. Although the estimated treatment effects are modest and marginally below conventional significance thresholds, their consistency across matching techniques supports the interpretation that selective disclosure strategies contribute to narrative misalignment. This evidence

reinforces arguments that fragmented ESG communication reflects strategic signaling choices rather than genuine sustainability engagement [19,28], highlighting the limits of relying solely on aggregate country-level ESG indicators to infer firm-level credibility.

4.1. Implications for Governance and Policy

From a governance and public policy perspective, the results advance the ESG and greenwashing debate by shifting attention from the quantity of disclosure toward the credibility and internal balance of ESG narratives. While existing EU sustainability regulation—including the Corporate Sustainability Reporting Directive (CSRD) and the European Sustainability Reporting Standards (ESRS)—prioritizes disclosure expansion and harmonization, the findings indicate that increased transparency alone does not eliminate strategic narrative behavior. Instead, firms may respond to regulatory pressure by selectively emphasizing favorable ESG dimensions while downplaying others, resulting in imbalanced and potentially misleading sustainability communication.

The persistence of moderate Greenwashing Severity Index (GSI) values across most Central and Eastern European countries suggests that greenwashing represents a systemic governance challenge rather than an isolated firm-level anomaly. This interpretation aligns with prior index-based assessments documenting widespread but moderate exaggeration of sustainability narratives across European firms [29]. Importantly, the results extend this literature by demonstrating that such patterns persist even under increasingly standardized reporting regimes, thereby questioning the assumption that disclosure harmonization alone ensures ESG credibility.

From a digitalization and governance standpoint, the findings highlight the dual role of digital tools. While digitalization enhances transparency and accelerates information diffusion, it simultaneously amplifies opportunities for narrative manipulation in the absence of digitally enabled oversight. Sustainability governance is therefore strengthened not through disclosure expansion per se, but through the institutional adoption of analytical tools capable of detecting disclosure asymmetries, sentiment imbalances, and strategic narrative reweighting. This interpretation is consistent with existing evidence showing that digital governance mechanisms can reduce information asymmetries and constrain opportunistic ESG reporting when embedded in supervisory processes [12,32].

The policy implications for EU institutions and national regulators are substantial. First, ESG supervision frameworks should increasingly integrate external information sources, such as media coverage and independent stakeholder narratives, as complementary inputs to corporate disclosures. Second, third-party verification mechanisms should be expanded beyond environmental claims to encompass social and governance dimensions, where narrative flexibility and greenwashing severity appear particularly pronounced. Third, continuous, technology-enabled monitoring—rather than reliance on periodic reporting cycles—should become a core element of ESG oversight, particularly as new disclosure requirements under the CSRD come into force.

Finally, the observed heterogeneity across sectors and firm sizes supports the case for risk-based and sector-sensitive regulatory approaches. Firms operating in sectors subject to stronger public accountability exhibit lower GSI values, suggesting that sustained scrutiny effectively constrains opportunistic disclosure practices. This implies that EU sustainability governance would benefit from differentiated oversight intensity, prioritizing sectors and business models characterized by higher informational asymmetries and reputational incentives for greenwashing.

Overall, the findings reinforce the view that credible ESG governance requires a shift from disclosure-centric regulation toward digitally enabled, data-driven supervisory models, capable of preserving the integrity of corporate sustainability reporting and sustainable finance within the evolving EU regulatory landscape.

4.2. Contributions, Limitations, and Future Research

In summary, the study offers three core contributions. First, it demonstrates that ESG credibility varies systematically across countries, industries, and firm sizes, reflecting governance capacity and stakeholder scrutiny rather than disclosure intensity alone. Second, it establishes firm-level disclosure imbalance as a key mechanism underlying greenwashing risk. Third, it introduces an ESG-integrated, NLP-based framework that advances both the theoretical and methodological treatment of greenwashing as a governance phenomenon rather than a marketing anomaly. Collectively, these findings extend legitimacy and signaling theories by showing that, in the digital era, more disclosure does not necessarily imply more credibility [25,30].

Several limitations should be acknowledged. The analysis relies on textual data from sustainability reports and media coverage, which capture communicative alignment rather than verified operational performance. Media narratives may themselves be biased or incomplete, and reliance on English-language sources may underrepresent local ESG discourse. The cross-sectional design further limits insight into dynamic responses to regulatory change. Future research could extend this framework in several directions. First, the analysis could be supplemented with a spatial dimension by integrating environmental Geographic Information System (GIS) data, allowing for a systematic comparison between heavily polluting firms and less environmentally intensive firms. Linking NLP-based greenwashing indicators with location-specific emissions, land use, or environmental exposure data would enable a more direct assessment of whether narrative misalignment is associated with observable environmental externalities. In addition, future work could explicitly examine heterogeneity between firms subject to third-party verification or external assurance and those without such mechanisms, thereby assessing whether independent oversight moderates greenwashing severity.

Second, while the current study focuses on English-language disclosures and media coverage to ensure comparability, extending the analysis to multilingual corpora would capture local sustainability narratives and reduce potential bias arising from language standardization. Applying multilingual transformer models or cross-lingual embeddings would allow future studies to incorporate country-specific discourse and regulatory contexts more accurately.

Third, future research could triangulate NLP-based indicators with objective environmental, social, and governance outcomes, such as emissions data, workplace safety records, governance enforcement actions, or supply-chain audits, to strengthen causal inference between communication strategies and underlying performance. Finally, adopting longitudinal designs would allow researchers to examine how greenwashing dynamics evolve in response to regulatory reforms, particularly under the implementation of the Corporate Sustainability Reporting Directive (CSRD), and whether enhanced disclosure requirements lead to substantive improvements or merely more sophisticated narrative alignment.

Despite these limitations, the present study provides a replicable and empirically grounded framework for assessing corporate greenwashing in emerging European contexts and offers a foundation for integrating textual analytics with spatial, institutional, and performance-based sustainability data in future research.

5. Conclusions

This study proposed and applied an NLP-based framework to assess the credibility of corporate ESG communication by measuring discrepancies between firms' sustainability disclosures and independent media narratives. By constructing a Greenwashing Severity Index (GSI), the analysis reframed greenwashing as an observable governance outcome rather than a purely perceptual or marketing-related phenomenon. The results

demonstrated that ESG credibility is determined not by disclosure intensity alone but by the internal balance of ESG communication and its consistency with external scrutiny.

Empirically, the findings showed that greenwashing across Central and Eastern Europe is moderate but systemic. Most firms engage extensively with ESG language, yet their disclosures frequently reflect selective emphasis across ESG dimensions rather than balanced, verifiable representation. This pattern is persistent across countries, industries, and firm sizes, indicating that greenwashing is embedded in prevailing reporting practices rather than driven by isolated misconduct. Importantly, firms that disproportionately emphasize individual ESG pillars exhibited higher greenwashing severity even after controlling for country- and industry-level factors, suggesting that disclosure asymmetry is a firm-level strategic choice rather than a structural inevitability.

A key contribution of this study is the evidence that institutional and regulatory strength does not automatically ensure the credibility of ESG disclosure. Countries and sectors with advanced regulatory frameworks or high aggregate ESG performance do not systematically exhibit lower greenwashing severity. Instead, disclosure credibility appears to be driven primarily by firm-level governance capacity, narrative balance across ESG dimensions, and the intensity of stakeholder scrutiny. This finding challenges the prevailing assumption that regulatory maturity alone guarantees substantive transparency and underscores the structural limitations of compliance-oriented ESG reporting regimes.

From a governance perspective, the study underscores the dual role of digitalization. While digital tools facilitate transparency and comparability, they also enable increasingly sophisticated narrative optimization when analytical oversight is weak. The results therefore point to a shift in regulatory emphasis: from expanding disclosure requirements toward strengthening interpretive and analytical oversight. Digital reporting infrastructures need to be complemented by data-driven monitoring instruments capable of identifying imbalance, exaggeration, and strategic reweighting across ESG dimensions.

These results have direct and actionable policy implications. First, ESG regulation should evolve from volume-oriented disclosure mandates toward consistency-based supervisory frameworks, in which the internal balance and coherence across ESG pillars constitute explicit criteria of reporting credibility. Second, regulatory authorities and standard-setters could institutionalize the use of automated text-analytics tools—such as discrepancy-based indices and narrative alignment metrics—within ESG review processes to systematically identify firms exhibiting persistent misalignment between disclosed and externally observed sustainability narratives. Third, the findings support extending mandatory third-party assurance beyond environmental indicators to encompass social and governance disclosures, particularly in sectors characterized by strong reputational incentives, elevated information asymmetries, and limited external scrutiny.

Notwithstanding its contributions, this study is subject to several limitations that point to promising directions for future research. First, the analysis is based on textual data and therefore captures communicative alignment rather than verified operational ESG performance, while external media narratives may themselves be selective or biased. Future research could address this limitation by triangulating NLP-based indicators with objective environmental, social, or governance outcomes, including emissions data, labor standards, or board-level governance metrics. Incorporating environmental GIS data would further enable comparisons between highly polluting firms and less environmentally intensive activities. In addition, explicitly examining the moderating role of third-party assurance could clarify whether external verification effectively constrains narrative manipulation. Extending the framework to multilingual corpora would allow a more accurate capture of locally embedded ESG discourse and mitigate biases associated with English-language standardization. Finally, longitudinal research designs would facilitate

the assessment of how greenwashing dynamics evolve in response to regulatory reforms, such as the implementation of the CSRD.

Overall, the study provides a replicable, empirically grounded framework for assessing ESG credibility in emerging European contexts. By operationalizing greenwashing as a multidimensional governance challenge rather than a narrow communication failure, it contributes to a more nuanced understanding of sustainability reporting in the digital era and offers practical tools for strengthening accountability as ESG regulation continues to expand.

Author Contributions: Conceptualization, A.A.D., E.M.M., A.-M.M., I.B. and S.-C.N.; methodology, E.M.M. and A.-M.M.; software, A.-M.M.; validation, A.A.D., A.-M.M., E.M.M., I.B. and S.-C.N.; formal analysis, A.A.D., A.-M.M., I.B. and E.M.M.; investigation, A.A.D., A.-M.M., E.M.M., I.B. and S.-C.N.; resources, A.A.D., A.-M.M., E.M.M. and I.B.; data curation, E.M.M. and A.-M.M.; writing—original draft preparation, I.B., A.A.D. and A.-M.M.; writing—review and editing, I.B., A.A.D., A.-M.M. and E.M.M.; visualization, I.B., A.-M.M. and E.M.M.; supervision, I.B. and A.A.D.; project administration, A.A.D.; funding acquisition, A.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the EU’s NextGenerationEU instrument through Romania’s National Recovery and Resilience Plan—Pillar III-C9-2022-I8, managed by the Ministry of Research, Innovation, and Digitalization, as part of the project titled “Accountable Governance and Responsible Innovation in Artificial Intelligence,” contract no. 760047/23.05.2023, code CF 158/15.11.2022. The article processing charge (APC) was also funded by this project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original dataset is available upon request to the corresponding author. Our dataset is formed only by public data.

Acknowledgments: The authors gratefully acknowledge the Business Economics Data Science Lab (BEDSL) at the Bucharest University of Economic Studies for research infrastructure and support. The study was conducted within the BEDSL lab and in the framework of the projects “Accountable Governance and Responsible Innovation in Artificial Intelligence” (Contract No. 760047/23.05.2023).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ESG	Environmental, Social, and Governance
AI	Artificial Intelligence
CEE	Central and Eastern Europe
CSR	Corporate Social Responsibility
EU	European Union
LDA	Latent Dirichlet Allocation
NLP	Natural Language Processing
TF-IDF	Term Frequency–Inverse Document Frequency

Appendix A

Appendix A.1. Greenwashing Dictionary—1140 Words

eco-friendly, green, renewable, non-renewable, low-impact, responsible, sustainable, ethical, unethical, responsibility, fair, just, equitable, correct, bio, organic, all natural, natural, eco-conscious, eco-system, eco-tourism, eco-label, ecocid, eco-village, eco-driving, eco-innovation, environmental destruction, environmental devastation, ecological collapse, habitat destruction, environmental catastrophe, biocide, climate-beneficial, climate-healing, climate-friendly, regenerative, carbon negative, net negative emissions, climate-positive, compostable, biodegradable, non-compostable, non-biodegradable, indestructible, environmentally aware, environmentally friendly, green-minded, sustainability-focused, earth-conscious, environmentally unaware, environmentally irresponsible, unsustainable, consumptive, polluting, wasteful, non-eco-friendly, pollution, no plastic, plastic free, non-plastic, sustainable packaging, eco-friendly packaging, biodegradable packaging, plastic-based, plastic-heavy, non-sustainable packaging, non-biodegradable packaging, single-use plastic, organic certification, plant-derived, plant-based, animal-based, non-transparent sourcing, sustainability, sustainably, sustaining, unsustainably, sustainable practices, transparent, true, truth, verifiable, unverifiable, verifiability, unverifiability, verified information, unverified information, verified, unverified, confirmed, unconfirmed, confirmed information, unconfirmed information, debunked, disputed, misinformation, false, untrue, disinformation, erroneous information, rumors, misleading, mislead, inaccurate, accurate, propaganda, doubtfulness, certain, certainty, uncertainty, uncertainty, ambiguity, ambiguous, skepticism, questionable, assurance, legal, illegal, clarity, certitude, trust, distrust, misgiving, misrepresent, misrepresentation, mislabel, misdirection, misuse, recycled, recycled, overly recycled, chemical-laden, cruelty-free, irresponsible, unaccountable, unreliable, neglectful, reckless, careless, incompetent, indifferent, feckless, imprudent, zero waste, renewable energy, solar power, wind energy, geothermal energy, hydropower, clean energy, energy-efficient, low-carbon, carbon-neutral, zero-emissions, green energy, sustainable agriculture, organic farming, permaculture, agroforestry, reforestation, afforestation, conservation, biodiversity, wildlife protection, habitat restoration, ecosystem services, natural capital, circular economy, closed-loop system, upcycling, reusable, reuse, repair, repurpose, reduce, minimalism, low-waste, plastic-free, zero-plastic, composting, vermiculture, biophilia, green building, leed certification, passive house, net-zero building, green roof, urban greening, green infrastructure, sustainable transport, public transport, cycling, walking, electric vehicles, hybrid vehicles, carpooling, ride-sharing, telecommuting, remote work, digital nomad, green jobs, green economy, ethical investing, socially responsible investing, impact investing, green bonds, carbon credits, carbon trading, emissions trading, carbon offsetting, carbon sequestration, blue carbon, green finance, green banking, sustainable finance, ethical banking, green procurement, sustainable sourcing, fair trade, ethical trade, local sourcing, farm-to-table, slow food, non-gmo, heirloom seeds, seed saving, community-supported agriculture, farmers' market, urban farming, vertical farming, hydroponics, aquaponics, agroecology, soil health, regenerative agriculture, holistic management, rotational grazing, cover cropping, no-till farming, polyculture, companion planting, integrated pest management, natural pest control, biological control, pollinator-friendly, bee-friendly, bird-friendly, wildlife corridor, marine protected area, ocean conservation, coral reef protection, mangrove restoration, wetland conservation, watershed management, sustainable fisheries, fishery certification, marine stewardship, sustainable seafood, aquaculture, shellfish farming, seaweed farming, algae biofuel, bioenergy, biomass energy, biogas, biochar, green chemistry, non-toxic, natural cleaning products, refillable, bulk buying, package-free, sustainable fashion, slow fashion, ethical fashion, fair fashion, secondhand, thrifted, vintage, upcycled

fashion, recycled materials, natural fibers, organic cotton, hemp, bamboo, linen, wool, silk, tencel, modal, recycled polyester, econyl, sustainable tourism, ecotourism, responsible travel, green travel, low-impact travel, carbon-neutral travel, sustainable accommodation, green hotel, eco-lodge, eco-resort, community-based tourism, cultural tourism, wildlife tourism, nature-based tourism, adventure tourism, sustainable events, green events, zero-waste events, plastic-free events, sustainable catering, plant-based diet, vegan, vegetarian, flexitarian, reducetarian, meatless monday, plant-forward, climate diet, local food, seasonal food, organic food, fair trade food, food sovereignty, food security, food justice, food waste reduction, food recovery, food rescue, food sharing, community fridge, food bank, soup kitchen, sustainable water, water conservation, water efficiency, rainwater harvesting, greywater recycling, water footprint, virtual water, food waste, toxic, non-greenwashing, fraud, eco-fraud, false advertising, misleading claims, deceptive marketing, green pr, green hushing, green noise, green smokescreen, green gloss, green bluster, green bait, green bait-and-switch, green hypocrisy, eco-scam, green deception, environmental hoax, fake eco-friendly, misleading green claims, false sustainability, green charade, green bluff, environmental misrepresentation, green cover-up, green exaggeration, green misdirection, green misrepresentation, green puffery, green spin, community, vague, astroturfing, fake grassroots, synthetic grassroots, artificial grassroots, counterfeit grassroots, false grassroots, orchestrated campaign, disguised campaign, orchestrated, fabricated support, fake public support, simulated grassroots, manufactured consensus, deceptive grassroots, pseudo-grassroots, sham grassroots, phony grassroots, staged support, contrived grassroots, bogus grassroots, feigned grassroots, spurious grassroots, recyclable, returnable, replenishable, rechargeable, reloadable, reclosable, reclaimable, multi-use, sin, guilty, guilty, error, mistake, offense, trespass, missdead, ecology, regeneration, regenerate, regenerating, regenerated, regenerative design, regenerative practices, regenerative systems, regenerative processes, regenerative capacity, regenerative potential, regenerative development, regenerative economy, regenerative culture, regenerative communities, regenerative landscapes, regenerative urbanism, regenerative architecture, regenerative engineering, regenerative technologies, regenerative solutions, regenerative principles, regenerative approaches, regenerative strategies, regenerative methods, regenerative techniques, regenerative restoration, regenerative remediation, regenerative rehabilitation, regenerative rewilding, regenerative conservation, regenerative stewardship, regenerative management, regenerative planning, regenerative construction, regenerative manufacturing, regenerative production, regenerative consumption, regenerative circularity, regenerative cycles, regenerative loops, regenerative feedback, regenerative dynamics, regenerative resilience, regenerative adaptability, regenerative evolution, regenerative transformation, regenerative transition, regenerative innovation, regenerative mindset, regenerative thinking, regenerative worldview, regenerative paradigm, regenerative framework, regenerative model, regenerative theory, regenerative science, regenerative ecology, regenerative biology, regenerative biomimicry, regenerative biophilia, regenerative permaculture, regenerative agroecology, regenerative agroforestry, regenerative silviculture, regenerative horticulture, regenerative aquaculture, regenerative marine, regenerative freshwater, regenerative terrestrial, regenerative ecosystem, regenerative biosphere, regenerative planet, regenerative earth, regenerative life, regenerative living, regenerative future, regenerative sustainability, regenerative environmentalism, regenerative activism, regenerative movement, regenerative revolution, regenerative renaissance, regenerative awakening, regenerative consciousness, regenerative ethics, regenerative values, regenerative philosophy, regenerative spirituality, anti-environmental, anti-environmentalism, anti-environmentalist, anti-conservation, anti-conservationist, anti-consumer, anti-consumerism, anti-consumerist, anti-development, anti-globalization, anti-litter, anti-littering, anti-smog, greenwashing, green sheen, greenhushing, greenshifting, greencrowding,

greenlighting, greenlabelling, greenrinsing, eco-labeling, false environmental claims, environmental skepticism, climate change denial, climate denial, climate denier, environmental degradation, unsustainable practices, environmental harm, environmental damage, polluter, greenwashing tactics, misleading language, vague terminology, unsubstantiated claims, deceptive terminology, misleading buzzwords, environmental misinformation, eco-deception, greenwashing practices, greenwashing strategies, greenwashing accusations, greenwashing allegations, greenwashing scandals, greenwashing controversies, greenwashing criticism, greenwashing backlash, greenwashing penalties, greenwashing fines, greenwashing lawsuits, greenwashing regulations, greenwashing guidelines, greenwashing oversight, greenwashing monitoring, greenwashing enforcement, greenwashing exposure, greenwashing awareness, greenwashing education, greenwashing prevention, greenwashing detection, greenwashing reporting, greenwashing transparency, greenwashing accountability, greenwashing responsibility, greenwashing impact, greenwashing consequences, greenwashing effects, greenwashing implications, greenwashing risks, greenwashing challenges, greenwashing issues, greenwashing problems, greenwashing solutions, greenwashing mitigation, greenwashing reduction, greenwashing elimination, greenwashing eradication, greenwashing control, greenwashing management, greenwashing policies, greenwashing measures, greenwashing actions, greenwashing initiatives, greenwashing efforts, greenwashing campaigns, greenwashing programs, greenwashing projects, greenwashing activities, greenwashing operations, greenwashing behaviors, greenwashing attitudes, greenwashing beliefs, greenwashing perceptions, greenwashing perspectives, greenwashing views, greenwashing opinions, greenwashing stances, greenwashing positions, greenwashing ideologies, greenwashing movements, greenwashing trends, greenwashing patterns, greenwashing dynamics, greenwashing factors, greenwashing drivers, greenwashing influences, greenwashing determinants, greenwashing causes, greenwashing origins, greenwashing sources, greenwashing roots, greenwashing history, greenwashing evolution, greenwashing development, greenwashing growth, greenwashing expansion, greenwashing spread, greenwashing proliferation, greenwashing prevalence, greenwashing frequency, greenwashing occurrence, greenwashing incidence, greenwashing rate, greenwashing level, greenwashing degree, greenwashing extent, greenwashing magnitude, greenwashing severity, greenwashing intensity, greenwashing effect, greenwashing consequence, greenwashing outcome, greenwashing result, greenwashing aftermath, greenwashing legacy, greenwashing footprint, greenwashing mark, greenwashing trace, greenwashing sign, greenwashing symptom, greenwashing indicator, greenwashing signal, greenwashing warning, greenwashing alert, greenwashing caution, greenwashing notice, greenwashing announcement, greenwashing declaration, greenwashing statement, greenwashing message, greenwashing communication, greenwashing discourse, greenwashing narrative, greenwashing rhetoric, greenwashing propaganda, greenwashing spin, greenwashing manipulation, greenwashing distortion, greenwashing fabrication, greenwashing falsification, greenwashing misrepresentation, greenwashing exaggeration, greenwashing overstatement, greenwashing understatement, greenwashing minimization, greenwashing trivialization, greenwashing dismissal, greenwashing denial, greenwashing rejection, greenwashing opposition, greenwashing resistance, greenwashing defiance, greenwashing challenge, greenwashing contestation, greenwashing dispute, greenwashing conflict, greenwashing confrontation, greenwashing clash, greenwashing battle, greenwashing fight, greenwashing struggle, greenwashing war, greenwashing campaign, greenwashing crusade, greenwashing mission, greenwashing quest, greenwashing journey, greenwashing adventure, greenwashing exploration, greenwashing discovery, greenwashing investigation, greenwashing inquiry, greenwashing research, greenwashing study, greenwashing analysis, greenwashing examination, greenwashing inspection, greenwashing review,

greenwashing assessment, greenwashing evaluation, greenwashing appraisal, greenwashing audit, greenwashing check, greenwashing test, greenwashing trial, greenwashing experiment, greenwashing pilot, greenwashing prototype, greenwashing model, greenwashing example, greenwashing case, greenwashing instance, greenwashing scenario, greenwashing situation, greenwashing condition, greenwashing context, greenwashing environment, greenwashing setting, greenwashing background, greenwashing landscape, greenwashing terrain, greenwashing field, greenwashing domain, greenwashing area, greenwashing region, greenwashing zone, greenwashing sector, greenwashing industry, greenwashing market, greenwashing economy, greenwashing society, greenwashing culture, greenwashing community, greenwashing group, greenwashing organization, greenwashing institution, greenwashing entity, greenwashing body, greenwashing agency, greenwashing department, greenwashing division, greenwashing unit, greenwashing team, greenwashing staff, greenwashing personnel, greenwashing workforce, greenwashing labor, greenwashing employment, greenwashing job, greenwashing occupation, greenwashing profession, greenwashing career, greenwashing trade, greenwashing business, greenwashing enterprise, greenwashing company, greenwashing corporation, greenwashing firm, greenwashing partnership, greenwashing venture, greenwashing startup, greenwashing project, greenwashing initiative, greenwashing program, greenwashing plan, greenwashing strategy, greenwashing policy, greenwashing measure, greenwashing action, greenwashing step, greenwashing move, greenwashing decision, greenwashing choice, greenwashing option, greenwashing alternative, greenwashing solution, greenwashing remedy, greenwashing fix, greenwashing cure, greenwashing treatment, greenwashing intervention, greenwashing regulation, greenwashing decision compliance, greenwashing adherence, greenwashing observance, greenwashing conformity, greenwashing alignment, greenwashing integration, greenwashing coordination, greenwashing cooperation, greenwashing collaboration, greenwashing alliance, greenwashing coalition, greenwashing network, greenwashing association, greenwashing federation, greenwashing union, greenwashing league, greenwashing confederation, greenwashing consortium, greenwashing syndicate, greenwashing cartel, greenwashing trust, greenwashing monopoly, greenwashing oligopoly, greenwashing duopoly, greenwashing rivalry, greenwashing contest, greenwashing match, greenwashing game, greenwashing sport, greenwashing race, greenwashing tournament, greenwashing championship, greenwashing series, greenwashing season, greenwashing event, greenwashing occasion, greenwashing happening, greenwashing incident, greenwashing episode, greenwashing phenomenon, greenwashing trend, greenwashing pattern, greenwashing cycle, greenwashing phase, greenwashing stage, promising, mismatch, camouflage, imprecise, vagueness, humanwashed, unsurprisingly, irrelevancy, unattended, muteness, oppositely, unrealistic, microplastic.

Appendix A.2. Environmental Dictionary—122 Words

Agroforestry, Air quality, Biodegradable materials, Biodiversity, Carbon footprint, Carbon management, Carbon neutrality, Carbon offsetting, Carbon sequestration, Circular economy, Clean air, Clean energy, Clean technology, Clean water, Climate action, Climate adaptation, Climate change, Climate policy, Climate resilience, Climate science, Coastal protection, Conservation, Deforestation, Eco-friendly, Ecological footprint, Ecological restoration, Ecosystem services, Ecotourism, Emission reduction, Emission trading, Emissions inventory, Energy conservation, Energy efficiency, Energy storage, Energy transition, Environmental advocacy, Environmental auditing, Environmental compliance, Environmental education, Environmental footprint, Environmental governance, Environmental health, Environmental impact assessment, Environmental innovation, Environmental justice, Environmental management, Environmental policy, Environmental

preservation, Environmental reporting, Environmental risk management, Environmental standards, Environmental stewardship, Forest management, Fossil fuels, Green bonds, Green buildings, Green certification, Green construction, Green finance, Green infrastructure, Green procurement, Green technology, Greenhouse gas emissions, Greenhouse gases, Habitat protection, Habitat restoration, Hazardous waste, Invasive species, Land conservation, Land degradation, Low carbon technologies, Low-carbon economy, Low-carbon, Low-emission vehicles, Marine biodiversity, Marine conservation, Natural capital, Natural resource conservation, Natural resource management, Ocean acidification, Organic farming, Plastic reduction, Pollution prevention, Recycling, Reforestation, Renewable energy, Renewable energy certificates, Renewable energy sources, Renewable resources, Resource efficiency, Resource recovery, Resource scarcity, Soil conservation, Soil health, Solar energy, Sustainability, Sustainable agriculture, Sustainable business practices, Sustainable design, Sustainable development, Sustainable development goals, Sustainable fisheries, Sustainable forestry, Sustainable innovation, Sustainable investment, Sustainable materials, Sustainable supply chains, Urban greening, Waste management, Waste-to-energy, Water conservation, Water efficiency, Water footprint, Water quality, Water scarcity, Water stewardship, Water sustainability, Wildlife conservation, Wildlife habitats, Wildlife protection, Zero waste.

Appendix A.3. Social Dictionary—122 Words

ABC, Access to education, Affordable housing, Anti bribery and corruption, Anti-discrimination, B Corps, Career development, Child labor, Civic responsibility, Code of conduct, Code of ethics, Community value, Community engagement, Community investment, Community partnerships, Community relations, Consumer protection, Corporate awareness of family life, Corporate citizenship, Corporate human rights, Corporate human rights due diligence, Corporate social responsibility, CSR, Cultural competence, Data privacy, DEI, Diversity and inclusion, “Diversity, Equity and Inclusion”, Education of employees, Employee empowerment, Employee engagement, Employee participation, Employee relations, Employee well-being, Employees representatives, Equal opportunities, Equal opportunity, Equal treatment, Ethical labor practices, Ethical recruiting, Ethical supply chains, Executive remuneration, Fair trade practices, Fair wages, Family leave policies, Forced labor, Gender balance, Gender equality, Gender gap, Gender pay gap, Gender pay gap reporting, Global reporting initiative, Global reporting initiative standards, GRI, GRI Standards, Health and safety, Human capital management, Human rights, Humanitarian aid, ILO Standards, Inclusive policies, Labor practices, Living wage, Mental health support, Modern slavery, NFRD, Non-discrimination policies, Non-financial reporting directive, Occupational health and safety, OHS, Parental entitlements, Philanthropy, Privacy rights, Race to the bottom, Rights of indigenous peoples, SASB, SASB standards, Social audits, Social clauses, Social clauses in supplier contracts, Social dialogue, Social finance, Social impact, Social inclusion, Social innovation, Social justice, Social performance, Social responsibility, Social responsibility standard, Social return on investment, Social risks, Social ROI, Social value, Social washing, Socially responsible Investing, SRI, Stakeholder advocacy, Stakeholder engagement, Supply chain transparency, Sustainability accounting standards board, Sustainability accounting standards board standards, Transparency in recruitment processes, Transparent hiring processes, UN Principles on business and human rights, Values-based investing, Volunteerism, Whistle-blowing, Workforce development, Workforce disclosure initiative, Work-life balance, Workplace diversity, Workplace safety.

Appendix A.4. Governance Dictionary—122 Words

Accountability, Anti-bribery, Anti-corruption, Anti-fraud policies, Anti-fraud, Anti-money laundering, Audit committee, Beneficial ownership, Board accountability, Board diversity, Board evaluation, Board independence, Board oversight, Board structure, Board tenure, Bribery prevention, Business continuity, Business ethics, Business integrity, Code of conduct, Code of ethics, Compliance, Compliance monitoring, Conflict of interest, Conflict resolution, Corporate bylaws, Corporate citizenship, Corporate culture, Corporate ethics, Corporate governance, Corporate governance code, Corporate policies, Corporate responsibility, Corporate social responsibility, Corporate transparency, Crisis management, Cyber risk, Cybersecurity, Data governance, Data privacy, Disclosure practices, Diversity and inclusion, Diversity policies, Due diligence, “Environmental, Social, and Governance integration”, ESG integration, Ethical business practices, Ethical decision-making, Ethical guidelines, Ethical leadership, Executive accountability, Executive compensation, Executive oversight, Fair disclosure, Fair labor practices, Fiduciary duty, Financial integrity, Fraud prevention, Governance audits, Governance best practices, Governance committee, Governance disclosure, Governance diversity, Governance framework, Governance metrics, Governance policies, Governance principles, Governance risk, Governance standards, Independent audit, Independent directors, Information security governance, Insider trading policies, Institutional integrity, Internal audit function, Internal controls, Investor relations, Investor rights, Leadership accountability, Leadership ethics, Legal accountability, Legal compliance, Legal governance, Legal risk, Management accountability, Management ethics, Management practices, Non-executive directors, Non-financial reporting, Organizational ethics, Policy enforcement, Proxy voting, Regulatory compliance, Regulatory oversight, Remuneration committee, Reporting standards, Responsible investing, Risk assessment, Risk governance, Risk management, Risk oversight, Shareholder activism, Shareholder engagement, Shareholder meetings, Social responsibility governance, Stakeholder engagement, Stakeholder management, Stakeholder rights, Strategic oversight, Succession planning, Supply chain governance, Sustainability governance, Sustainability reporting, Tax transparency, Transparency, Transparency in lobbying, Transparency in political donations, Transparency initiatives, Transparent reporting, Values-based governance, Voting rights, Whistleblower protection.

References

1. Kathan, M.C.; Utz, S.; Dorfleitner, G.; Eckberg, J.; Chmel, L. What you see is not what you get: ESG scores and greenwashing risk. *Financ. Res. Lett.* **2025**, *74*, 106710. <https://doi.org/10.1016/j.frl.2024.106710>.
2. Dorfleitner, G.; Utz, S. Green, green, it's green they say: A conceptual framework for measuring greenwashing on firm level. *Rev. Manag. Sci.* **2024**, *18*, 3463–3486. <https://doi.org/10.1007/s11846-023-00718-w>.
3. Lublóy, Á.; Keresztúri, J.L.; Berlinger, E. Quantifying firm-level greenwashing: A systematic literature review. *J. Environ. Manag.* **2025**, *373*, 123399. <https://doi.org/10.1016/j.jenvman.2024.123399>.
4. Spaniol, M.J.; Danilova-Jensen, E.; Nielsen, M.; Rosdahl, C.G.; Schmidt, C.J. Defining Greenwashing: A Concept Analysis. *Sustainability* **2024**, *16*, 9055. <https://doi.org/10.3390/su16209055>.
5. Saxena, S. Using natural language processing for detecting greenwashing indicators and constructing impact-focused index portfolio. *SSRN Electron. J.* **2024**. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5024113 (accessed on 18 November 2024).
6. Ziolo, M.; Bąk, I.; Spoz, A. Literature Review of Greenwashing Research: State of the Art. *Corp. Soc. Responsib. Environ. Manag.* **2024**, *31*, 5343–5356. <https://doi.org/10.1002/csr.2842>.
7. Jeevan, P. Green Washing—A Conceptual Framework. In *Development Prospects of Indian Economy*; Bharathi Publications: Delhi, India, 2014; pp. 270–276, ISBN 978-93-81212-62-2.
8. Yoganandham, G.; Kareem, A.A.; Khan, E.M.I. Unveiling the Shadows—Corporate Greenwashing and Its Multifaceted Impacts on Environment, Society, and Governance: A Macroeconomic Theoretical Assessment. *Shanlax Int. J. Arts Sci. Humanit.* **2024**, *11*, 20–29. <https://doi.org/10.34293/sijash.v11iS3-Feb.7237>.

9. Cude, B.J. Consumer Perceptions of Environmental Marketing Claims: An Exploratory Study. *J. Consum. Stud. Home Econ.* **1993**, *17*, 207–225. <https://doi.org/10.1111/j.1470-6431.1993.tb00166.x>.
10. Lim, W.M.; Ting, D.H.; Bonaventure, V.S.; Sendiawan, A.P.; Tanusina, P.P. What Happens When Consumers Realise About Green Washing? A Qualitative Investigation. *Int. J. Glob. Environ. Issues* **2013**, *13*, 14–24. <https://doi.org/10.1504/IJGENVI.2013.057323>.
11. Mangini, E.R.; Amaral, L.M.; Conejero, M.A.; Pires, C.S. Greenwashing Study and Consumers' Behavioral Intentions. *Rev. Bras. Mark. Negócios* **2020**, *4*, 229–244. <https://doi.org/10.51359/2526-7884.2020.244488>.
12. Pham, N.T.; Barretta, P.G. Green Marketing or Greenwashing: How Consumers Evaluate Environmental Ads. *J. Appl. Bus. Econ.* **2024**, *26*, 19. <https://doi.org/10.33423/jabe.v26i1.6808>.
13. Dahl, R. Green Washing: Do You Know What You're Buying? *Environ. Health Perspect.* **2010**, *118*, A246–A252. <https://doi.org/10.1289/ehp.118-a246>.
14. Fella, S.; Bausa, E. Green or Greenwashed? Examining Consumers' Ability to Identify Greenwashing. *J. Environ. Psychol.* **2024**, *95*, 102281. <https://doi.org/10.1016/j.jenvp.2024.102281>.
15. Sajid, M.; Zakkariya, K.A.; Suki, N.M.; Islam, J.U. When Going Green Goes Wrong: The Effects of Greenwashing on Brand Avoidance and Negative Word-of-Mouth. *J. Retail. Consum. Serv.* **2024**, *78*, 103773. <https://doi.org/10.1016/j.jretconser.2024.103773>.
16. Xu, T.; Sun, Y.; He, W. Government Digitalization and Corporate Greenwashing. *J. Clean. Prod.* **2024**, *452*, 142015. <https://doi.org/10.1016/j.jclepro.2024.142015>.
17. Bernini, F.; Giuliani, M.; La Rosa, F. Measuring Greenwashing: A Systematic Methodological Literature Review. *Bus. Ethics Environ. Responsib.* **2024**, *33*, 649–667. <https://doi.org/10.1111/beer.12631>.
18. Schmidt, S.; Kinne, J.; Lautenbach, S.; Blaschke, T.; Lenz, D.; Resch, B. Greenwashing in the US Metal Industry? A Novel Approach Combining SO₂ Concentrations from Satellite Data, a Plant-Level Firm Database and Web Text Mining. *Sci. Total Environ.* **2022**, *835*, 155512. <https://doi.org/10.1016/j.scitotenv.2022.155512>.
19. Kim, J.S.; Sim, J.B.; Kim, Y.J.; Park, M.K.; Oh, S.J.; Doo, I.C. Establishment of NLP-Based Greenwashing Pattern Detection Service. In *Advances in Computer Science and Ubiquitous Computing*; Park, J.S., Yang, L.T., Pan, Y., Park, J.H., Eds.; Lecture Notes in Electrical Engineering; Springer: Singapore, 2023; Volume 1028. https://doi.org/10.1007/978-981-99-1252-0_32.
20. Chang, K.; Shim, H.; Yi, T.D. Corporate Social Responsibility, Media Freedom, and Firm Value. *Financ. Res. Lett.* **2019**, *30*, 1–7. <https://doi.org/10.1016/j.frl.2019.03.019>.
21. Venturelli, A.; Caputo, F.; Leopizzi, R.; Pizzi, S. The State of Art of Corporate Social Disclosure Before the Introduction of the Non-Financial Reporting Directive: A Cross-Country Analysis. *Soc. Responsib. J.* **2019**, *15*, 409–423. <https://doi.org/10.1108/SRJ-12-2017-0275>.
22. Linnenluecke, M.K. Environmental, Social and Governance (ESG) Performance in the Context of Multinational Business Research. *Multinatl. Bus. Rev.* **2022**, *30*, 1–16. <https://doi.org/10.1108/MBR-11-2021-0148>.
23. Arvidsson, S.; Dumay, J. Corporate ESG Reporting Quantity, Quality and Performance: Where to Now for Environmental Policy and Practice? *Bus. Strategy Environ.* **2022**, *31*, 1091–1110. <https://doi.org/10.1002/bse.2937>.
24. Li, M.; Chen, Q. Executive Pay Gap and Corporate ESG Greenwashing: Evidence from China. *Int. Rev. Financ. Anal.* **2024**, *95*, 103375. <https://doi.org/10.1016/j.irfa.2024.103375>.
25. Zhang, D. The Pathway to Curb Greenwashing in Sustainable Growth: The Role of Artificial Intelligence. *Energy Econ.* **2024**, *133*, 107562. <https://doi.org/10.1016/j.eneco.2024.107562>.
26. Rusu, R.A.; Manea, D.I.; Paraschiv, D.M. Quantifying Sustainability: An Empirical Analysis of ESG Integration in CEE Pension Funds. *Proc. Int. Conf. Bus. Excell.* **2025**, *19*, 3460–3471. <https://doi.org/10.2478/picbe-2025-0264>.
27. Bancu, A. A Meta-Analysis of ESG Disclosure and Company's Economic Performance. *Proc. Int. Conf. Bus. Excell.* **2024**, *18*, 2042–2056. <https://doi.org/10.2478/picbe-2024-0173>.
28. Ito, S.; Itoi, T.; Năstase, M.; Nicolae, B. ESG Effect on the Corporate Value of Technology Companies. *Proc. Int. Conf. Bus. Excell.* **2024**, *18*, 1929–1940. <https://doi.org/10.2478/picbe-2024-0162>.
29. De Lucia, C.; Pazienza, P.; Bartlett, M. Does Good ESG Lead to Better Financial Performances by Firms? Machine Learning and Logistic Regression Models of Public Enterprises in Europe. *Sustainability* **2020**, *12*, 5317. <https://doi.org/10.3390/su12135317>.
30. Kalaitzoglou, I.; Pan, H.; Niklewski, J. Corporate Social Responsibility: How Much Is Enough? A Higher Dimension Perspective of the Relationship Between Financial and Social Performance. *Ann. Oper. Res.* **2021**, *306*, 209–245. <https://doi.org/10.1007/s10479-020-03834-y>.

31. Domanović, V. The Relationship between ESG and Financial Performance Indicators in the Public Sector: Empirical Evidence from the Republic of Serbia. *Manag. J. Sustain. Bus. Manag. Solut. Emerg. Econ.* **2022**, *27*, 69–80. <https://doi.org/10.7595/management.fon.2021.0032>.
32. Lagasio, V. Measuring Greenwashing: The Greenwashing Severity Index. *SSRN Electron. J.* **2023**. <https://doi.org/10.2139/ssrn.4582917>.
33. Schimanski, T.; Reding, A.; Reding, N.; Bingler, J.; Kraus, M.; Leippold, M. Bridging the Gap in ESG Measurement: Using NLP to Quantify Environmental, Social, and Governance Communication. *Financ. Res. Lett.* **2024**, *61*, 104979. <https://doi.org/10.1016/j.frl.2024.104979>.
34. Zhao, Y.; Kroher, L.; Engler, M.; Schnattinger, K. Detecting Greenwashing in the Environmental, Social, and Governance Domains Using Natural Language Processing. In Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2023), Rome, Italy, 13–15 November 2023; KDIR; SCITEPRESS—Science and Technology Publications: Setúbal, Portugal, 2023; Volume 1, pp. 175–181. <https://doi.org/10.5220/0012155400003598>.
35. Yin, L.; Yang, Y. How Does Digital Finance Influence Corporate Greenwashing Behavior? *Int. Rev. Econ. Financ.* **2024**, *93*, 359–373. <https://doi.org/10.1016/j.iref.2024.05.014>.
36. Coface. *CEE Top 500: Insights on Central Europe's Largest Firms*; Coface: Bucharest, Romania, 2024. Available online: <https://www.coface.ro/en/news-economy-expert-advice/coface-cee-top-500-ranking-leading-companies-in-a-challenging-economic-landscape> (accessed on 19 November 2024).
37. Muhammad Abdullah. *Gnews*, Version 0.4.2.; Python Package. Available online: <https://pypi.org/project/gnews/> (accessed on 18 November 2024).
38. Pearson, K. Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia. *Philos. Trans. R. Soc. Lond. Ser. A* **1896**, *187*, 253–318.
39. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
40. Spärck Jones, K. A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *J. Doc.* **1972**, *28*, 11–21. <https://doi.org/10.1108/eb026526>.
41. Refinitiv. *Refinitiv ESG Scores: Methodology*; Refinitiv (LSEG Data & Analytics): London, UK, 2023. Available online: https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf (accessed on 18 November 2024).
42. Austin, P.C. Optimal Caliper Widths for Propensity-Score Matching When Estimating Differences in Means and Differences in Proportions in Observational Studies. *Pharm. Stat.* **2011**, *10*, 150–161. <https://doi.org/10.1002/pst.433>.
43. Ho, D.E.; Imai, K.; King, G.; Stuart, E.A. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Anal.* **2007**, *15*, 199–236. <https://doi.org/10.1093/pan/mpl013>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.