Sustainable Futures 10 (2025) 101494

Contents lists available at ScienceDirect

Sustainable Futures

journal homepage: www.sciencedirect.com/journal/sustainable-futures

Large language models for sustainability reporting: A systematic review and research agenda

Seyed Alireza Mousavian Anaraki [1], Danilo Croce *,[1], Roberto Basili [1]

Department of Enterprise Engineering, University of Rome Tor Vergata, Via del Politecnico 1, Rome, 00133, Italy

ARTICLE INFO

Keywords:
Sustainability reporting
Information extraction
Greenwashing
Large language models
Meta-synthesis
Scientometric

ABSTRACT

Sustainability reporting refers to the activities by which large companies describe their efforts to maximize the benefits on universal criteria about economic, social, and environmental impacts. This promotes transparency among stakeholders and is considered essential in supporting governance and compliance with international regulations by improving today's business landscape. Among other problems, inaccurate company disclosures, known as *Greenwashing*, can affect the quality of reports and their unstructured nature poses serious limitations to economic analysts during company evaluation. Efficiently gathering and aligning available reports and other complementary data for a unified framework is an important perspective on modern sustainability analysis. Recent advances in Natural Language Processing (NLP), particularly the rise of transformer-based Large Language Models (LLMs), have enabled new capabilities in semantic information extraction, automated classification, and the detection of misleading claims within unstructured corporate sustainability disclosures. This survey provides a systematic review of NLP and LLM-based approaches in sustainability reporting, with particular attention to methodological trends and unresolved challenges. Adopting a structured review methodology that integrates scientometric mapping and meta-synthesis, we offer an evidence-based taxonomy and a synthesis of current practices, highlighting key gaps and future research directions. The findings highlight past explorations and future directions, demonstrating that LLMs offer remarkable accuracy and innovative solutions for challenges related to traditional sustainability reporting current practices.

1. Introduction

Corporate sustainability is a key concept in today's business world, focusing on operations that ensure long-term success, which entails sustainability reporting that considers economic, social, and environmental impacts to enhance transparency for stakeholders [1]. Sustainability reporting covers various types of reports like Corporate Social Responsibility (CSR), Greenhouse Gas Emissions (GHG), and Triple Bottom Line (TBL) reporting [2].

In the realm of sustainability reporting, the United Nations (UN) Agenda 2030 [3] and Global Reporting Initiative (GRI)[2] are vital because they provide comprehensive frameworks and standards for organizations to measure, report, and improve their economic, social, and environmental performance. The 2030 Agenda outlined 17 Sustainable Development Goals (SDGs) in 2015 to be achieved by 2030, while the GRI sets a structured approach and offers standards to report the economic, environmental, and social impacts of sustainability practices [4]. Additionally, financial access is crucial in corporate sustainability [5], and ensuring true sustainability is essential for investors and responsible capital distribution [6]. Companies adhering to the Environmental, Social, and Governance (ESG) framework stand a better chance of securing financing. While there's no single ESG framework for corporate decision-making, multiple standards exist based on industry, location, stakeholder expectations, and organizational priorities [7]. Despite its advantages, corporate sustainability faces challenges, such as efficiently organizing diverse data for CSR insights [8] and combating "Greenwashing", where companies falsely portray sustainable practices to mislead consumers and regulators for a clean image [9]. Some researchers restrict the term greenwashing to environmental issues while using "Bluewashing" to address social concerns. In contrast, other scholars regard greenwashing as a broader concept that includes both social and environmental dimensions [10].

* Corresponding author.
E-mail addresses: seyedalireza.mousaviananaraki@students.uniroma2.eu (S.A. Mousavian Anaraki), croce@info.uniroma2.it (D. Croce), basili@info.uniroma2.it (R. Basili).

[1] All authors contributed equally to this work.

[2] https://www.globalreporting.org/standards/

https://doi.org/10.1016/j.sftr.2025.101494
Received 30 December 2024; Received in revised form 4 October 2025; Accepted 1 November 2025
Available online 21 November 2025
2666-1888/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

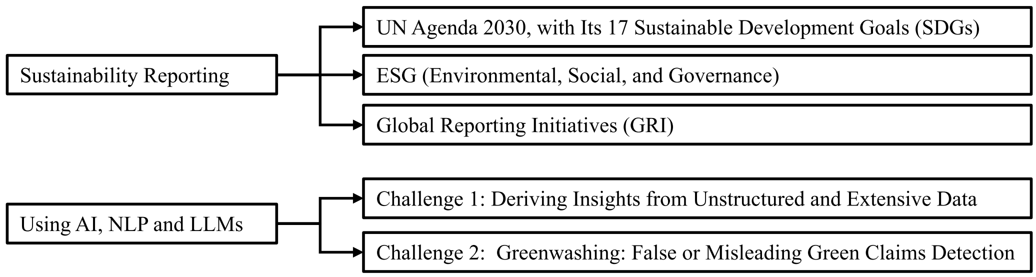

Fig. 1. Review Road Map.

Sustainability reporting is increasingly complex, with the volume of available information growing rapidly, a first major challenge for organizations [11]. In addition, extracting actionable insights and detecting misleading green claims present further difficulties [12]. Natural Language Processing (NLP), as a core area of Artificial Intelligence (AI), offers significant potential to enhance both the precision and efficiency of these tasks [13]. In particular, Large Language Models (LLMs) have demonstrated remarkable accuracy in a wide range of NLP applications [14,15] and can provide innovative solutions for validating company claims through the analysis of sustainability reports. These models have proven effective in improving data extraction, report analysis, and green claim detection, making them highly suitable for addressing the current challenges in sustainability reporting [16].

Although there is extensive academic literature on sustainability reporting [17,18], greenwashing [10], and LLMs [14,19], only one paper has systematically reviewed the intersection of all three areas through a bibliometric literature review up to 2023 [2]. However, this study did not include LLMs as part of the state-of-the-art advancements in NLP.

More recent contributions such as [20,21] have begun exploring the implications of generative AI and greenwashing detection, respectively, within sustainability contexts. De Villiers et al. [20] proposes a conceptual framework analyzing the potential of AI-generated text in sustainability accounting and reporting, while Boedijanto and Delina [21] highlights AI's role in detecting greenwashing in the energy sector. While both offer valuable conceptual insights, their scope remains limited: Boedijanto and Delina [21] focuses exclusively on the energy sector, and De Villiers et al. [20] concentrates primarily on sustainability accounting and assurance, rather than addressing the broader spectrum of sustainability reporting across diverse domains. Furthermore, neither study conducts a systematic literature review nor employs advanced scientometric techniques to quantitatively assess the existing research landscape, as this study does.

To date, no comprehensive systematic review has addressed the intersection of sustainability reporting, NLP, and LLMs. This study addresses this gap by employing a hybrid scientometric and meta-synthesis methodology and pursuing two overarching objectives: (1) extracting structured insights from unstructured sustainability reports, and (2) detecting misleading green claims (greenwashing) using NLP and LLMs. The review is grounded in the analysis of three major sustainability reporting frameworks (SDG, GRI, and ESG), which serve as reference points throughout the study.

As illustrated in Fig. 1, this review is organized around two central axes. The first axis concerns the main frameworks used for sustainability reporting, namely the UN Agenda 2030 with its SDGs, ESG, and GRI. The second axis addresses the main challenges that motivate the review's objectives, which arise in leveraging advanced NLP and LLM-based techniques for sustainability reporting, specifically: (1) the difficulty of deriving insights from unstructured and extensive data, and (2) the detection of false or misleading green claims (greenwashing). The intersection of these two perspectives forms the analytical basis of this review, allowing for a structured exploration of how recent language technologies contribute to addressing persistent challenges in sustainability disclosure, while accounting for the requirements and distinct features of each reporting framework.

The paper is structured as follows: Section 2 provides a comprehensive background on sustainability reporting, greenwashing detection, and LLMs by emphasizing the role of LLMs in enhancing them. Section 3 details the review design and methodological framework. Section 4 presents the findings and thematic synthesis derived from the selected studies. Section 5 outlines practical implications and future research directions. Finally, Section 6 draws the conclusions.

2. Background

Sustainability reporting involves encouraging organizations to reevaluate their strategies and operations, considering their economic, social, and environmental impacts, and to communicate this information to stakeholders who demand greater transparency on these issues [17]. In this context, GRI and SDG[3] are two major standards and frameworks in the global sustainability landscape. The GRI has recognized the importance of the SDGs and has worked to align its standards with the UN goals. While both aim to promote sustainability and corporate responsibility, they have slightly different approaches and purposes. Although SDGs and the GRI are significant frameworks for sustainability, they have encountered various challenges and criticisms. With a growing focus on sustainability, investors are increasingly influenced by GRI-based reporting, and several efforts have been carried out to achieve the SDGs [22]. Despite its broad applicability, the GRI may need to be tailored to fit the unique needs of specific regions or industries [23]. Additionally, potential conflicts between GRI standard and local regulations can create difficulties for organizations trying to adhere to both sets of standards [24]. In response to these challenges, ESG framework has emerged as an alternative, closely linked to financial outcomes [25]. An extended discussion of these frameworks is presented in Appendix A.

According to the Oxford English Dictionary,[4] "*greenwashing*" is defined as a deliberate corporate practice involving misleading elements aimed at deceiving stakeholders. Greenwashing is also defined as false or exaggerated claims about poor environmental performance to appear more responsible or sustainable to create a misleading positive image for the public and stakeholders [26]. Due to the expansion of sustainability today to include ESG factors, greenwashing now involves more than just environmental claims (as the most common form). It also encompasses misleading information about a company's social [27] and governance practices [10].

In sustainability reporting, greenwashing poses a significant threat to the credibility and transparency of corporate disclosures. At its core, greenwashing involves discrepancies between what a company chooses to disclose or signal to stakeholders and its actual environmental or social performance. This dishonesty not only tarnishes the company's reputation but also undermines broader efforts towards genuine sustainability [2]. When misleading green claims are detected, they erode

[3] https://sdgs.un.org/

[4] https://www.oed.com/

the trust of consumers and other stakeholders. This lost trust is challenging to be regained and can severely damage the company's image, potentially hindering its ability to attract investments and support in the future [28]. A more detailed classification of greenwashing forms and examples is provided in Appendix B.

In light of these challenges, particularly the widespread presence of unverifiable or misleading sustainability claims, there is a growing demand for automated, scalable techniques to analyze unstructured corporate disclosures and support the identification of potential greenwashing [7]. Traditional manual assessments are often time-consuming, costly, and subject to inconsistencies, especially when applied to lengthy sustainability reports across multiple frameworks [29].

Recent advances in NLP, particularly the development of LLMs, have opened promising avenues for addressing these limitations. LLMs, built upon the transformer architecture [30], are pre-trained on massive text corpora and exhibit strong capabilities in understanding, classifying, and generating natural language at scale [14]. Their ability to capture deep contextual relationships within text makes them particularly well-suited for tasks such as information extraction, report classification, and greenwashing detection.

The use of LLMs can be broadly grouped into three main architectural families, each offering distinct strengths in the context of sustainability reporting tasks [31]:

- **Encoder-Only (EO)** models, such as BERT [32], RoBERTa [33], and ClimateBERT [12], focus exclusively on understanding input text. They are particularly effective for classification tasks, such as aligning report content with GRI or SDG frameworks, and for detecting misleading claims embedded in corporate disclosures.
- **Decoder-Only (DO)** models, including the GPT family [34,35] and LLaMA [36], are optimized for generative tasks. These models can produce fluent, context-aware narratives and summaries, and are increasingly applied to generate or verify sustainability-related statements, especially when combined with Retrieval-Augmented Generation (RAG) [7].
- **Encoder-Decoder (ED)** models, such as T5 [37] and BART [38], integrate both comprehension and generation capabilities. They are particularly valuable for tasks like SDG multi-label classification, report summarization, or question-answering over ESG content [39].

These architectural differences also influence practical deployment. For instance, EO models are commonly preferred for analyzing short, paragraph-level content, while DO and ED models are often used when working with larger input sequences, such as full-page or multi-section documents [40].

Beyond classification and text generation, LLMs have also proven effective in identifying deceptive or exaggerated claims associated with greenwashing. For example, ClimateBERT, a fine-tuned EO model, has been used to extract structured insights from sustainability reports and flag potentially misleading statements [16]. While LLMs do not eliminate the need for expert validation, they drastically reduce the manual effort required for initial screening and provide a scalable foundation for more rigorous auditing.

Ultimately, LLMs enable the extraction of actionable insights from complex and voluminous text, help standardize evaluations across different frameworks (GRI, SDG, ESG), and assist in the early detection of inconsistencies that might indicate greenwashing. Their integration into sustainability reporting workflows thus represents a key step towards more transparent, efficient, and accountable corporate sustainability practices.

3. Review design and methodological framework

A robust and transparent review methodology is essential to synthesize advances at the intersection of sustainability reporting and language technologies. To this end, the present review adopts a structured, multi-step process, as summarized in Fig. 2, to ensure conceptual clarity, reproducibility, and methodological rigor.

The study is anchored in three interconnected **Research Dimensions**: (i) corporate sustainability reporting and its guiding frameworks (e.g., GRI, SDG, ESG); (ii) applications of NLP and LLMs; and (iii) the identification and analysis of greenwashing practices.

To operationalize the **Objectives**, we adopt a **Systematic Methodology** that integrates meta-synthesis principles [41,42] with scientometric analysis [43]. The process unfolds across seven key steps:

1. **Determining research questions** focused on information extraction and greenwashing detection.
2. **Systematic literature review** using major academic databases.
3. **Searching and selecting appropriate articles** based on defined inclusion and exclusion criteria.
4. **Data extraction from the articles** on frameworks used, techniques applied, and addressed challenges.
5. **Article analysis by scientometric tools** to map trends and thematic clusters.
6. **Quality control** using a standardized appraisal checklist.
7. **Findings presentation and qualitative combining** through classification and thematic grouping.

This mixed-method approach enables both a qualitative and quantitative examination of the literature, guided by two sets of criteria: selection criteria to ensure article relevance and rigor, and analytical criteria (such as sustainability framework, methodological scope, and greenwashing focus) to structure the synthesis.

The following outlines the seven steps of the systematic review methodology adopted in this study.

3.1. Determining research questions

In the initial stages of any structured literature review, the first task is to define the study's objectives and questions.

This paper focuses on addressing challenges in sustainability reporting, specifically:

Challenge 1: Extracting information and insights from unstructured data

Challenge 2: Combating greenwashing

These challenges are outlined in the introduction and are explored within the context of three sustainability frameworks: SDG, GRI, and ESG. The paper seeks to identify appropriate solutions presented in scientific literature by conducting a systematic literature review. It categorizes the most relevant papers based on the sustainability frameworks and the NLP and LLMs applications proposed for these challenges, ultimately providing directions for future research.

This study aims to answer the following questions:

*Q*1: "*What are the most impactful and influential research contributions at the intersection of sustainability reporting, NLP, and LLMs that can support future research and practice?*"

*Q*2: "*What are the main challenges and future directions in the field of sustainability reporting enhanced by NLP and LLMs?*"

3.2. Systematic literature review

The review process commenced with a systematic exploration of the Web of Science[5] (WoS) and Scopus[6] databases to identify relevant literature. A set of predefined logical queries was employed to retrieve studies at the intersection of sustainability reporting, NLP, and LLMs.

[5] https://www.webofscience.com/

[6] https://www.scopus.com

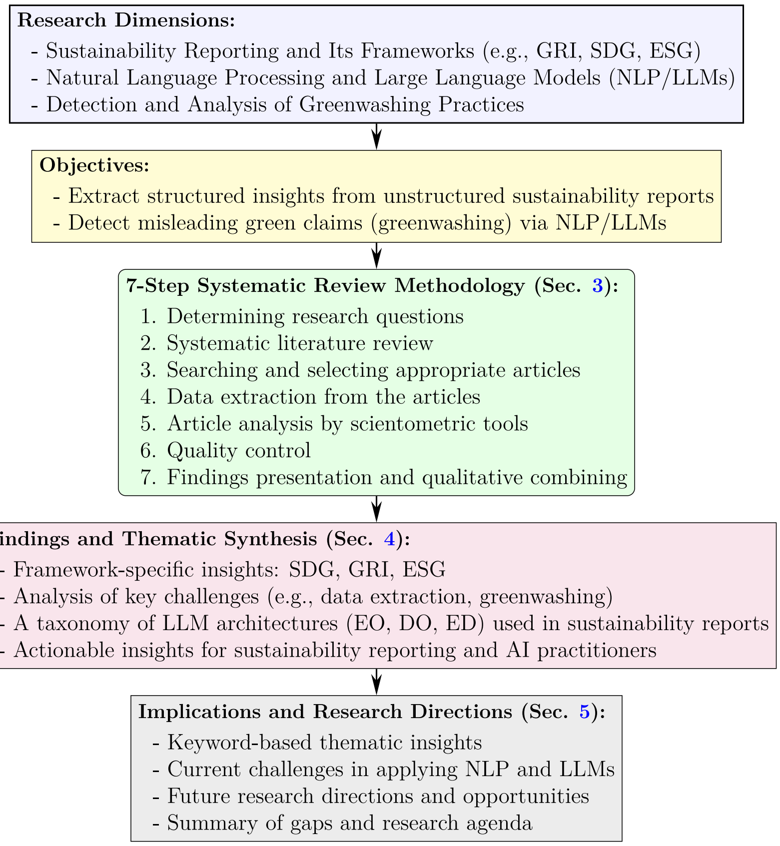

Fig. 2. Schematic overview of the review. The study builds on three research domains to address two objectives through a seven-step methodology. This leads to structured findings and future research directions.

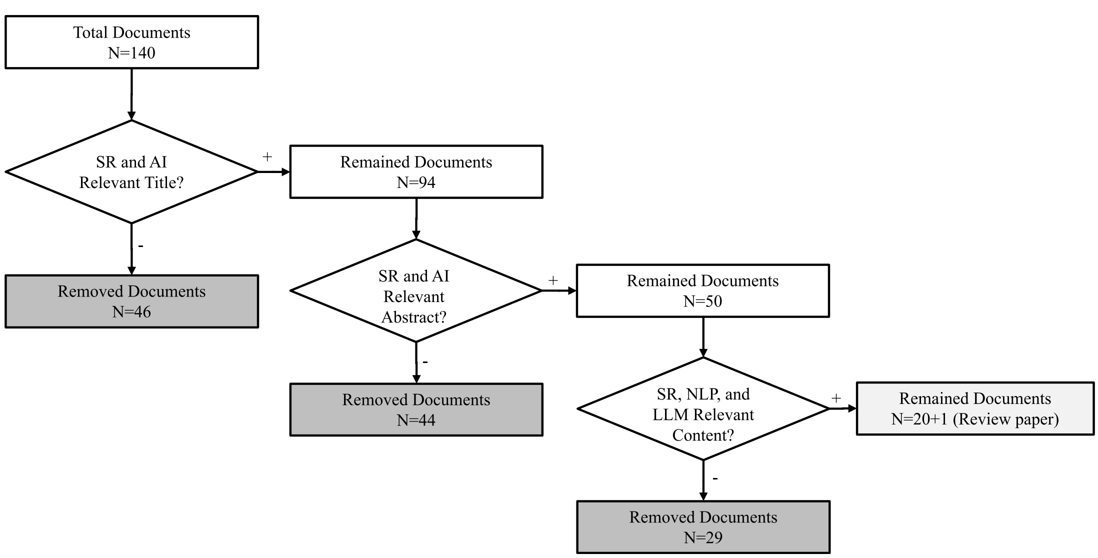

Fig. 3. Literature survey flowchart of selecting appropriate articles.

The exact queries used for each database are reported in Appendix C. The initial search yielded 68 records from WoS and 89 from Scopus. After merging results and removing duplicates using Mendeley,[7] 136 unique records were retained. An additional four papers were identified from sources such as Google Scholar, ResearchGate, and arXiv, resulting in a final corpus of 140 documents used for the review.

[7] https://www.mendeley.com

3.3. Searching and selecting relevant articles

Following the literature retrieval phase, a multi-stage screening was carried out to refine the initial dataset and retain only the most relevant studies. The process, illustrated in Fig. 3, involved a stepwise filtering of 140 documents.

In the first stage, titles were reviewed for relevance to sustainability reporting (SR) and AI-related techniques, excluding 46 papers. The remaining 94 abstracts were then screened, leading to the removal of

44 more. In the final stage, full-text reviews were conducted on 50 papers to assess alignment with NLP and LLM applications in SR. This resulted in the exclusion of 29 documents. The final selection comprised 21 articles, 20 original research papers, and one review article.

The inclusion criteria guiding this process were: (*i*) English-language publications from 2017 to 2024, (*ii*) direct relevance to SR, NLP, and LLMs, and (*iii*) empirical, methodological, or review-based contributions. The exclusion criteria were designed to ensure that only papers directly aligned with the focus of this review were retained. Specifically, we excluded documents such as editorials, opinion pieces, or technical notes that lacked substantive methodological or empirical contributions. We also excluded studies that addressed sustainability topics within narrow domains, such as agriculture, healthcare, or energy, when these did not explicitly involve sustainability reporting practices.

Further, papers relying solely on traditional analytical approaches (e.g., readability indices, complexity metrics, or key performance indicators (KPIs) dashboards) were omitted if they did not incorporate advanced NLP or LLM techniques. Basic forms of text mining, such as simple frequency analysis or topic modeling, were also excluded unless embedded in a broader NLP pipeline. Additionally, we discarded studies that mentioned NLP or LLMs only superficially or speculatively, without actual application to the analysis, classification, or interpretation of sustainability reports. For example, general AI applications to sustainability assessments were deemed out of scope unless directly tied to the content of corporate or institutional reporting documents. A categorized overview of excluded works and justification examples is provided in Appendix D.

It is important to acknowledge that the primary screening and selection were conducted by one reviewer. To mitigate potential bias, a random subset of papers was independently reviewed by a second author. The inter-rater agreement yielded a Cohen's κ of 0.85, indicating substantial agreement and enhancing the credibility of the selection process.

3.4. Data extraction from the articles

To facilitate a structured synthesis of the reviewed literature, we extracted key information from each of the 21 selected studies. Table 1 summarizes this information for the 20 original research articles (excluding the single review paper). Each entry has been classified along four dimensions:

- **Framework:** the sustainability reporting standard addressed in the study (e.g., GRI, ESG, SDG);
- **Methodology:** the NLP models and techniques applied, including both traditional and transformer-based LLMs;
- **Challenge 1 (Ch. 1):** whether the study tackles the challenge of extracting insights from unstructured sustainability-related data;
- **Challenge 2 (Ch. 2):** whether the study addresses greenwashing detection or the identification of misleading sustainability claims.

Articles are sorted chronologically to reflect the methodological evolution in the field, from early NLP applications to the adoption of cutting-edge transformer architectures and generative models. This classification provides the foundation for the thematic synthesis presented hereafter.

Table 1
Extracted information from the articles, sorted by year and methodology evolution. The "Framework" column indicates the sustainability framework used (e.g., GRI, ESG, SDG). The "Methodology" column lists the NLP models and techniques employed in the study. Columns "Ch. 1" and "Ch. 2" refer to the two key challenges addressed in the respective studies, with an asterisk (*) denoting that the challenge was covered and a dash (-) indicating it was not.
Source: Research Finding

Reference	Framework	Methodology	Ch. 1	Ch. 2
[44]	GRI	NLP, GloVe	*	–
[11]	GRI	NLP, IR	*	–
[45]	ESG	LinkBERT, FinBERT	*	*
[46]	SDG	MiniLM, DistilBERT	*	–
[1]	SDG	BERT, RoBERTa, MLs	*	–
[5]	ESG	SBERT-UN, RoBERTa	*	*
[47]	ESG	ClimateBERT	*	*
[4]	GRI	TF-IDF, RF	*	–
[16]	ESG	BERT	*	*
[48]	GRI	BERT	*	–
[12]	ESG	ClimateBERT	*	*
[49]	ESG	MPNet	*	–
[39]	SDG	BART, BERT, GPT	*	–
[7]	ESG/GRI	GPT, ChatGLM, RAG	*	–
[50]	ESG	NLP	*	–
[25]	ESG	BERT, YAKE	*	–
[51]	SDG	RoBERTa	*	–
[52]	ESG	RoBERTa, DistilRoBERTa	*	–
[40]	SDG	GPT, Longformer	*	–
[8]	ESG	T5, LLaMA, RAG	*	–

3.5. Article analysis by scientometric tools

Building upon the structured data extraction presented in Table 1, this step focuses on identifying overarching patterns, research clusters, and thematic directions through scientometric analysis. Rather than providing additional article-level summaries, the aim here is to interpret the scientific landscape across sustainability frameworks by exploring how NLP and LLMs have been used to address key challenges in sustainability reporting. To this end, we applied two complementary tools widely adopted in bibliometric research. VOSviewer (v1.6.20) was used to construct and visualize keyword co-occurrence networks, highlighting dominant topics and interlinked subthemes. In parallel, Bibliometrix (R package, v4.3.1) enabled the extraction of descriptive and structural bibliometric indicators. Together, these tools support both qualitative insights and quantitative trend analysis, helping to position individual studies within broader thematic trajectories. The analysis revealed three primary areas of focus, corresponding to the main sustainability frameworks considered in this review: the SDG, GRI, and ESG standards. We therefore organized the discussion of findings around these clusters, discussing how different NLP and LLM techniques were employed in each domain to extract insights from unstructured sustainability reports and detect or mitigate greenwashing. In the remaining, first, we discuss the articles grouped by their framework alignment (SDG, ESG, GRI), identifying common methodological approaches and contributions. Then, we provide a synthesis of the review papers included in our dataset, focusing on how they frame the current state and gaps of NLP-driven sustainability research.

SDG. The analysis of sustainability reports through NLP and LLMs, particularly in relation to SDGs, has been a growing area of research. Building on this, Kang and Kim [46] proposed new methods, including sentence similarity techniques and sentiment analysis to study sustainability reports using pre-trained MiniLM-BERT for sentence similarity and DistilBERT for sentiment analysis on sustainability reports from six global leader companies from 2011 to 2020. Results confirmed that companies use sustainability reports to enhance their positive image during crises. However, the importance of continuous monitoring for report reliability and temporal analyses was emphasized. In a broader context, Angin et al. [1] aimed to examine various Machine Learning (ML) approaches composed of Support Vector Machines (SVM), Decision Tree (DT), Logistic Regression (LR), and Gaussian Naive Bayes (GNB) optimized for NLP-based text classification tasks to classify reports according to their relevance to the SDGs. They fine-tuned the Robustly Optimized BERT Approach (RoBERTa) model for processing sustainability reports on the Open source SDG Community Dataset

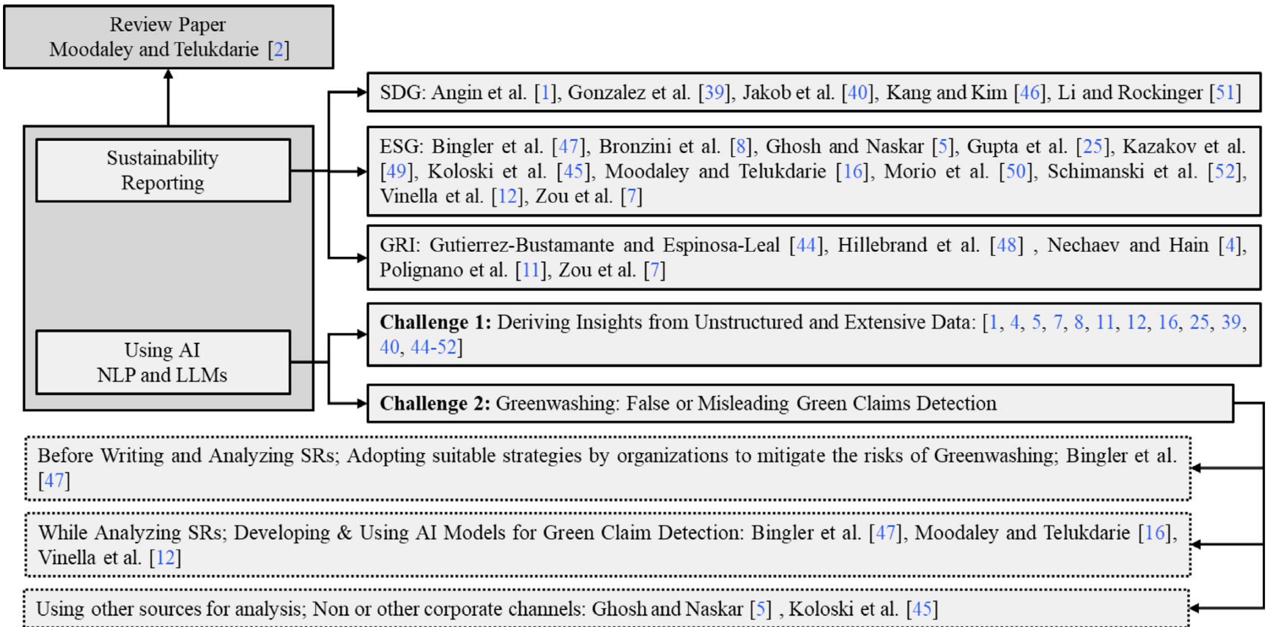

Fig. 4. Qualitative combining findings.

(OSDG-CD),[8] which contains textual data on the SDGs annotated by community volunteers. The results demonstrate that the fine-tuned RoBERTa model achieved very high performance in both binary and multi-class classification tasks. However, the study is limited to the OSDG-CD and suggests future work to classify national AI strategy documents of over 50 countries. Gonzalez et al. [39] aimed to analyze NLP for Social Good (NLP4SG) papers and introduce a novel dataset and system to support this analysis. The work presented the PaperAnalyzer system, consisting of three LLM-based models to perform binary classification of NLP4SG papers, multi-class categorization into the 17 UN SDGs, and identification of salient scientific terms. SciBERT+DS performs best for binary classification, InstructGPT excels in multi-label categorization compared to BART and DistilBERT, and InstructGPT-002 achieves the highest BERTScore for task and method extraction. However, the study is limited to knowledge within NLP, suggesting future work should collect data on the downstream real-world impact of research to provide better feedback. Li and Rockinger [51] analyzed the thematic evolution of sustainability reports of 98 listed European banks from 2010 to 2022 by developing a novel RoBERTa model for classification, incorporating word counts and k-means cluster analysis to track changes in banking objectives. The study showed effective differentiation of texts across various SDGs with high accuracy (93%). The research also identified strong correlations between SDG prevalence and macroeconomic indicators. Finally, Jakob et al. [40] investigated the impact of incorporating companies' self-assessments on classifying SDG contributions in sustainability reports. The methodology involved fine-tuning transformer-based models (OpenAI-GPT, Transformer-XL, Longformer) and training a CNN to automate labeling sustainability reports due to the time-consuming nature of manual labeling. The best results were obtained using Longformer. However, the study emphasizes the need for guidelines and labeling of sustainability reports as a fundamental step. It suggests that future work should focus on developing guidelines and further streamlining the labeling process.

ESG. The application of NLP and LLMs to ESG analysis has gained considerable attention, with several researchers exploring how best to classify and assess financial texts and sustainability reports. Koloski et al. [45] focused on the FinSim4-ESG task, where they classified sentences from financial reports as either sustainable or unsustainable. They employed two representation paradigms: text-based and knowledge-based. The researchers fine-tuned two variants of BERT: LinkBERT and FinBERT, to improve the classification of ESG sentences. Their future work will include training deep neural networks based on sentence representations and conducting feature importance analysis to enhance classification performance. Similarly, Ghosh and Naskar [5] aimed to classify financial texts as sustainable or unsustainable and rank ESG-related concepts for unknown terms. They fine-tuned multiple models, including a Sentence BERT-based model pre-trained on UNSDG (SBERT-UN) for ranking ESG concepts, and a RoBERTa model for text classification. FinBERT performed best on the validation set, while RoBERTa-base excelled on the test set. However, it is limited to the specific models and datasets used, and further improvements could include more diverse data sources and advanced NLP techniques. The research also suggested future work in developing a user-friendly tool for assigning terms to concepts and evaluating the sustainability of financial texts. Bingler et al. [47] took a more targeted approach by analyzing companies' climate-risk disclosures within the Task Force on Climate-Related Financial Disclosures (TCFD) framework using a specifically designed deep neural language model fine-tuned with BERT called ClimateBERT. Training on a dataset consisting of over 17,300 human-labeled sentences for the TCFD pillars and more than 300,000 general language sentences from annual reports. However, the study highlights the issue of greenwashing and suggests that further work is needed to enhance the model's ability to detect genuine climate risk information. Addressing the issue of greenwashing more directly, Moodaley and Telukdarie [16] proposed a conceptual framework built on BERT-based architectures that use subdomain-specific text corpora. Their framework effectively detected false or misleading green claims, though limitations included the use of a potentially limited training dataset and BERT's inherent word associations. They suggested expanding the dataset and refining the model for more accurate detection in future work. Following this theme, Vinella et al. [12] presented a new method for training a language model to identify greenwashing risk using the ClimateBERT model. The ClimateBERT model was

[8] https://github.com/osdg-ai/osdg-data

pre-trained on 2 million climate-related text samples and further fine-tuned on sustainability reports. The model demonstrated an average accuracy of 86.34%. The approach involved a mathematical framework for quantifying greenwashing risk. However, some limitations of the approach include potential errors in the mathematical formulations, skewed label distribution, and the loss of context due to the paragraph-by-paragraph analysis. The researchers plan to refine the model and address these limitations in future work. Due to the limitations of earlier models like BERT and XLNet, Kazakov et al. [49] introduced "ESGify" an open-source NLP model based on MPNet-base architecture, aimed at classifying texts within ESG risk frameworks. MPNet is an innovative model that combines Masked Language Modeling (MLM) and Permuted Language Modeling (PLM) and addresses the limitations of previous models. Although it has been successful, there is a need for further work to refine the model and broaden its applicability. Zou et al. [7] presented ESGReveal, an innovative method for extracting and analyzing ESG data from sustainability reports, addressing the lack of a comprehensive ESG disclosure database. They used LLMs including GPT-4. QWEN, GPT-3.5, ChatGLM combined with RAG techniques on ESG reports from 166 companies listed on the Hong Kong Stock Exchange. The system achieved an accuracy of 76.9% in data extraction and 83.7% in disclosure analysis. The study calls for further research to develop and compare the analytical capabilities of various LLMs, enhancing ESG data processing and analysis. The tool known as ReportParse, introduced in [50], is a Python-based application engineered to dissect corporate sustainability reports. The tool seamlessly integrates document structure analysis with NLP models to effectively extract pertinent sustainability-related data from the reports. Gupta et al. [25] aimed to examine the sustainability practices of seven leading IT firms in India by analyzing their 2021 sustainability reports, focusing on the balance across ESG factors. They employed BERT-based tokenization and fine-tuning BERT for sequence classification combined with the Unsupervised Automatic Keyword Extraction (YAKE) technique. They achieved a high prediction accuracy of 98%. However, they could add reports in a wider spectrum of sectors and conduct a time-series study. They also suggested considering more current dimensions, such as sustainable human resource management and employee well-being, in future research. Schimanski et al. [52] addressed the gap in precise, transparent ESG measurement by using pre-trained RoBERTa and DistilRoBERTa models for each ESG subdomain. They achieved over 93% accuracy for social and environmental domains and over 89% for governance. The models' performance suggested effective ESG communication detection, but further investigations are needed to explore patterns beyond correlations. Bronzini et al. [8] used advanced NLP techniques to extract insights on ESG factors from sustainability reports. They employed Instruction-tuned LLMs, In-Context Learning (ICL), and RAG paradigm to analyze data from various sources: (*i*) 2022 sustainability reports, (*ii*) 64 ESG factors, and (*iii*) ESG rating scores and other companies' information. However, the study highlighted the challenge of extracting meaning from diverse and unstructured documentation, suggesting future integration of additional data sources and improved information extraction techniques.

GRI. Similar to the previous two frameworks and standards, special attention has also been given to the GRI in research, so that: Gutierrez-Bustamante and Espinosa-Leal [44] aimed to analyze Nordic companies' alignment with GRI standard and automate sustainability report reviews for environmental impact. They employed NLP and Information Retrieval (IR) methods to obtain semantic similarity on 524 sustainability reports from the GRI standard dataset. The models they used included Bag of Words (BOW), Skip-gram, Term Frequency–Inverse Document Frequency (TF–IDF), Latent Semantic Analysis (LSA), Word2Vec, FastText, Global Vectors for Word Representation (GloVe). Their findings identified LSA and GloVe as the most effective methods. However, they noted issues with the quality of the data and recommended enriching the dataset with a wider range of frameworks such as ESG and SDGs for future work. Polignano et al. [11] presented a system based on NLP and IR techniques to automatically analyze sustainability reports compliant with GRI standards. They identified various sustainability topics, determined their context, and assessed whether they were mentioned in a positive or negative light. This approach significantly accelerated the analysis of corporate documents. Nonetheless, future endeavors will focus on improving the system's robustness, efficiency, and flexibility. Nechaev and Hain [4] contributed to the sustainability reports literature by employing a pre-trained language model for identifying social impact types aligning GRI codes. They used a novel Standard-based Impact Classification method (SBIC) that employs TF–IDF for vectorization and Random Forest (RF) for classification. The study found a positive correlation between sustainability reports and innovation capacity. Limitations included the need for more advanced ML techniques and improved text preprocessing accuracy. Future research should explore multi-label classification and state-of-the-art embedding techniques. Hillebrand et al. [48] introduced sustain.AI, an advanced recommender system leveraging two German sustainability reports datasets: GRI (92 reports from major companies) and DNK (German Sustainability Code database). The system employs a BERT-based encoding module and multi-label classification to recommend relevant text passages aligned with GRI standards. It has performed better than other systems in recommending segments. However, it could only handle German documents. Future enhancements included incorporating user feedback for model improvement and extending functionality to process English reports. The work of [7] that was mentioned in the previous (ESG) part, used three main standards composed of TCFD, GRI, and Sustainability Accounting Standards Board (SASB) as ESG frameworks.

Review Paper. Among the studies examined, only one review article was included in the final corpus: [2]. This work presents a comprehensive conceptual framework at the intersection of AI, ML, greenwashing, and sustainability reporting. It employs both bibliometric and thematic analyses to identify key patterns and emerging topics within this area. Although the review does not explicitly focus on LLMs, it lays a valuable foundation for understanding the broader methodological context in which LLMs are beginning to emerge. As such, it complements the empirical studies analyzed in this review and helps identify gaps for future research. A summary of additional review papers excluded from the final analysis is provided in Appendix D.

3.6. Quality control

To ensure the validity and reliability of the findings derived from the selected literature, a formal quality assessment was conducted as part of the meta-synthesis process. We employed the Critical Appraisal Skills Programme (CASP)[9] framework, one of the most widely used tools for evaluating the rigor, relevance, and transparency of research [53]. CASP provides specialized checklists tailored to different study types, including qualitative research, quantitative analyses, and systematic reviews.

Each paper was evaluated using a ten-question CASP checklist, adapted to the specific nature and methodological approach of the study. The questions covered aspects such as clarity of aims, appropriateness of methodology, data analysis quality, and ethical considerations. Responses were categorized as "*Yes*", "*No*", or "*Can't tell*", following CASP guidelines. The first two items served as screening questions; only papers receiving a "*Yes*" on both were subjected to full evaluation.

This structured assessment aimed to reduce bias and enhance the consistency of the appraisal process. In line with CASP recommendations, no aggregate score was calculated; instead, the distribution of "*Yes*" responses was used as an indicator of overall quality.

[9] https://casp-uk.net/casp-tools-checklists/

The results confirmed the robustness of the selected corpus:

- **90%** of papers received "*Yes*" for all applicable criteria.
- The remaining **10%** received "*Yes*" on at least 90% of the items, with minor limitations noted, typically related to the Researcher–Participant Relationship or Accuracy of Data Analysis.

This process ensured that only high-quality, methodologically sound studies were included in the final synthesis. The full CASP-based evaluation for each article is presented in Appendix E.

3.7. Findings presentation and qualitative synthesis

The integration of findings from the selected articles marks a pivotal step in this review, forming the foundation for all subsequent analysis, discussion, and research directions. Rather than reporting outcomes from individual studies, this synthesis distills broader insights, recurring patterns, and emerging trends that define the intersection of sustainability reporting, NLP, and LLMs. By connecting and interpreting the evidence gathered in earlier steps, this qualitative synthesis provides a comprehensive overview of the current landscape, guiding the critical reflections and future agenda addressed in the following sections.

Fig. 4 visually categorizes the 21 reviewed articles into two main groups: a single review paper and 20 research papers. The research articles are further classified by their alignment with major sustainability frameworks: SDG, GRI, and ESG, as well as the specific challenges addressed through NLP and LLM approaches.

- **Framework Alignment:** Most research papers target ESG (11), followed by SDGs (5) and GRI (5). This reflects the rising regulatory and investor-driven importance of ESG, while SDG and GRI remain crucial but slightly less dominant focal points for recent AI research.
- **Applications for Challenge 1-Extracting Insights from Unstructured Data:** All 20 research papers address the primary challenge of extracting information and insights from unstructured or extensive sustainability data. This is typically approached as a classification or structured extraction problem. Early studies predominantly relied on EO models such as BERT, RoBERTa, and their domain-specific variants. More recent work explores ED and DO architectures, including generative models like GPT-3, particularly for tasks involving summarization or data transformation.
- **Applications for Challenge 2-Greenwashing Detection:** A smaller but significant subset of papers confront the challenge of detecting false or misleading green claims (greenwashing). The literature reveals three main strategies:

    1. *Pre-reporting risk mitigation:* Some papers focus on best practices and strategies organizations can adopt before reports are written to minimize greenwashing risk, such as establishing governance structures and transparent policies [47].
    2. *In-report detection:* Other studies apply NLP and LLMs directly to the analysis of sustainability reports, leveraging fine-tuned models like ClimateBERT or domain-adapted BERT variants to identify misleading or exaggerated claims [12,16,47].
    3. *External cross-validation:* A third group extends analysis by integrating additional sources such as financial news or regulatory documents, enabling models to cross-reference corporate claims with independent data [5,45].

  This multi-layered approach demonstrates how the field is advancing from isolated text analysis towards ecosystem-aware, multi-source validation of sustainability claims.

Fig. 5 presents a taxonomy that maps these studies not only by sustainability framework but also by the neural architectures adopted-EO, ED, and DO, in relation to the two core challenges. EO models remain foundational, particularly for classification-heavy tasks, due to their efficiency in mapping complex corporate text into analyzable formats. Both ED and DO models are gaining ground, especially for more generative or interpretative applications (e.g., automated summaries, complex information extraction).

Notably, there is an emerging trend towards the adoption of DO models such as GPT-3, reflecting a growing interest in generative capabilities for report summarization, interpretative assistance, and the creation of training data for downstream classification. At the same time, the literature continues to emphasize the power of domain-adapted EO model, like ClimateBERT and FinBERT, for highly specialized classification and detection tasks.

In summary, while EO architectures currently underpin most sustainability reporting applications, especially those centered on text classification, there is clear evidence of diversification. The adoption of generative and hybrid architectures is enabling more sophisticated, context-aware, and scalable solutions, both for extracting actionable insights and for tackling complex phenomena like greenwashing.

This dual synthesis serves as a springboard for the next two sections: we now move from mapping the field to a critical evaluation of the most impactful research contributions (Section 4) and, subsequently, to an exploration of open challenges and future research priorities (Section 5).

4. Findings and thematic synthesis

This section synthesizes the main findings emerging from the reviewed literature, building directly on the classification and analysis conducted in the previous steps. Rather than reiterating individual study outcomes, the goal here is to consolidate broader thematic insights and methodological patterns.

The analysis confirms that most contributions concentrate on two recurring challenges: the extraction of structured information from unstructured sustainability reports, and the identification of misleading or overstated claims. Across the reviewed works, a progressive adoption of increasingly advanced NLP models is evident, ranging from early encoder-based architectures to more recent generative approaches. The studies also vary in their alignment with sustainability frameworks such as ESG, SDGs, and GRI, reflecting different priorities and use cases.

This synthesis provides a structured foundation for the critical reflections and future research directions that follow, where the implications of these findings are further examined in relation to open challenges, methodological gaps, and opportunities for advancement in the field.

Response to the research Q1. By interpreting the results from the Table 1 and Fig. 4, we can find that all 20 related papers, except one review paper [2], on sustainability reporting, NLP, and LLMs aimed to derive insights from sustainability reports addressing *Challenge 1*: "*Extracting information and insights from unstructured data*". The sustainability frameworks considered in these papers varied: five papers focused on SDG, five papers focused on GRI and 11 papers focused on ESG.

Among these, some papers also addressed *Challenge 2*: "*Combating greenwashing*", and their methodologies for solving the second challenge can be divided into three main approaches:

1. Adopting suitable strategies by organizations to mitigate the risks and negative impacts of greenwashing [47] and promote transparency, integrity, and accountability, including:

    - **Strong Governance:** Implementing clear policies and procedures helps prevent greenwashing by providing a framework for accountability and transparency.

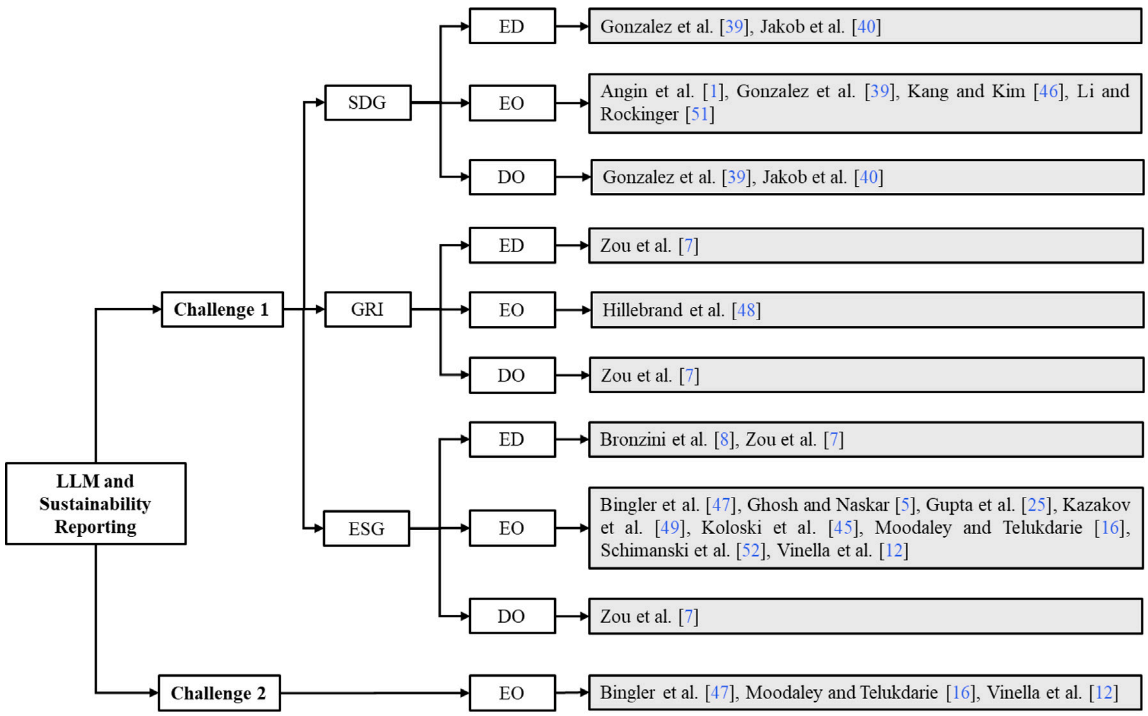

Fig. 5. Papers categorized by their alignment with SDG, GRI, and ESG frameworks, and the types of neural models applied: ED, EO, and DO.

- **Full and Accurate Disclosure:** Disclosing complete and truthful information about performance and practices allows consumers and stakeholders to make informed decisions, thus preventing greenwashing.
- **Due Diligence:** Ensuring that sustainability claims are accurate and verifiable reduces the risk of greenwashing.
- **Clear Communication:** Effectively communicating a company's sustainability strategy mitigates the risks associated with greenwashing by ensuring that all stakeholders properly understand the claims.
- **Education and Training:** Educating and training staff about what constitutes greenwashing and how to avoid it can prevent these misleading practices within the organization.
- **Collaboration with Third Parties:** Working with independent third parties to verify sustainability claims enhances transparency and helps prevent greenwashing.
- **Stakeholder Engagement:** Actively listening to and engaging with stakeholders provides valuable feedback and helps companies align their sustainability strategies with stakeholder expectations.

2. While analyzing sustainability reports; developing and using LLMs for green claim detection [12,16,47].
3. Using non or other corporate channels; Since sustainability reports are created by companies and are considered company-controlled sources, some studies utilize non-corporate channels such as business and financial news [54], media coverage data [55], but analyzing media coverage is challenging due to the increasing number of daily news articles, the need to filter relevant topics from media coverage data carefully, and the predominantly unstructured nature of the data [56]. So, some papers used other corporate channels [5,45] and integrated reports [57].

Regarding the adopted methods and neural architectures, the analysis also shows the taxonomy of papers in terms of Sustainability Reporting-based and LLM structured-based taxonomy, enabling future researchers to easily identify which LLM methods and structures are used to address specific challenges in sustainability reporting. By interpreting the results from Table 1 and Fig. 5, we can find: Gonzalez et al. [39] by using BART (ED), BERT (EO), GPT (DO) considering all three LLM structures, Angin et al. [1],Kang and Kim [46], and Li and Rockinger [51] by using BERT, SBERT, RoBERTa, MiniLM, DistilBERT (EO structure), and finally Jakob et al. [40] by using OpenAI-GPT, Transformer-XL, and Longformer (ED and DO structures) tried to solve first challenge for sustainability reporting considering SDG. It can be interpreted in the same way, for other papers so that for GRI, Hillebrand et al. [48] used BERT (EO structure), and Zou et al. [7] used LLMs combined RAG (ED and DO structures). For ESG, Zou et al. [7],Bronzini et al. [8] used LLMs combined RAG (ED and DO structures). Bingler et al. [47], Ghosh and Naskar [5], Gupta et al. [25], Kazakov et al. [49], Koloski et al. [45], Moodaley and Telukdarie [16], Schimanski et al. [52], and Vinella et al. [12] used ClimateBERT, SBERT-UN, RoBERTa, BERT, YAKE, MPNet, LinkBERT, FinBERT, DistilRoBERTa (EO structure).

Among these papers, some papers also considered the second challenge so that Bingler et al. [47], Moodaley and Telukdarie [16], and Vinella et al. [12] used ClimateBERT and BERT (EO structure) for green claim detection. It should be noted some papers such as [4,11,44,50] weren't in this taxonomy because they focused on NLP-based methods.

Overall, it is worth noting the dominant use of EO models like BERT and RoBERTa, which have proven highly effective for the classification tasks central to sustainability reporting. Their fine-tuning capabilities allow for extracting relevant insights from unstructured data, as demonstrated in studies focused on aligning corporate reports with frameworks like the SDGs, GRI, and ESG.

However, there is an increasing shift towards DO models, particularly in the context of generative tasks. The growing adoption of models like GPT-3 highlights the demand for more sophisticated tools capable of summarization and generating insights from complex reports. These models aid in producing reader-friendly summaries, helping to transform how sustainability reports are interpreted and consumed by both technical and general audiences. DO models also

Table 2
LLMs addressing key challenges in sustainability reporting tasks.

LLM-based contribution	Architecture	Frameworks
Unstructured Data Extraction: Extracting structured information from unstructured reports through automatic classification and topic mapping	EO, DO, ED	SDG, GRI, ESG
Cross-Framework Alignment: Linking concepts across SDG, GRI, and ESG standards using semantic similarity	EO, ED	SDG, GRI, ESG
Data Preprocessing and Filtering: Prompting and generating high-quality labeled data to improve classification performance	DO, ED	SDG, GRI, ESG
Report Summarization: Producing concise and accessible summaries of lengthy sustainability disclosures	DO, ED	GRI, ESG
Greenwashing Detection: Identifying misleading claims using fine-tuned domain-specific models	EO	ESG

EO = Encoder-Only, DO = Decoder-Only, ED = Encoder-Decoder. Source: Author's synthesis.

serve as effective prompting and filtering mechanisms, guiding the extraction of relevant information and curating high-quality labeled data for downstream classification tasks performed by EO models. This pre-processing capability helps improve the performance and reliability of EO-based pipelines. This trend suggests that the use of DO models may expand, especially for tasks that require more dynamic interaction with sustainability data or for improving the quality of inputs used in analytical models.

Models of ED, while used less frequently, have shown their value in tasks that combine understanding and generation, such as summarizing reports. They bridge the gap between comprehension and content creation, making sustainability disclosures more accessible to both experts and lay readers. Future developments may involve hybrid architectures that combine EO and DO features, potentially leading to more accurate and adaptable solutions for both extracting insights and generating new content in sustainability reporting.

Finally, future research may focus on specialized models, like ClimateBERT, or on multi-modal data integration-leveraging not only textual content but also visuals, structured data, and external media. This direction could enhance the holistic assessment of sustainability performance, enabling better detection of greenwashing and aligning with the growing demand for corporate transparency and accountability.

To improve readability and provide a concise synthesis of the main findings, Table 2 presents a structured summary of how LLMs contribute to key tasks in sustainability reporting. Each row captures a specific LLM-enabled functionality, ranging from information extraction to generative tasks, highlighting the types of neural architectures employed and the sustainability frameworks addressed. This overview emphasizes the breadth of LLM applications across the sustainability reporting pipeline. Models of EO, such as BERT and RoBERTa, dominate classification-related tasks, including unstructured data extraction and claim verification. Both DO and ED models are more frequently applied to tasks requiring generation, such as report summarization and data enrichment through prompting. These generative models are also used to improve downstream classification performance by supplying higher-quality training data.

Importantly, all model types engage with multiple sustainability frameworks, including SDGs, GRI, and ESG, underlining the importance of cross-framework adaptability in real-world applications. The observed distribution suggests an emerging trend towards hybrid strategies that integrate classification and generation to address the growing complexity and interpretability needs of sustainability reporting.

Taken together, these insights offer a consolidated understanding of how LLMs contribute to overcoming key analytical and operational challenges in sustainability reporting, setting the stage for identifying the emerging research needs and future opportunities in this evolving field.

5. Implications and research directions

Building on the thematic synthesis presented earlier, this section explores the broader implications of current research at the intersection of sustainability reporting, NLP, and LLMs. We aim to identify key challenges that remain unresolved and outline future directions to guide both scholarly inquiry and practical innovation in the field. To support this analysis, we examine the conceptual structure emerging from the reviewed literature through a co-occurrence network of keywords. This allows us to map the intellectual landscape of the domain and highlight clusters of concepts that represent core topics, methodological trends, and underexplored areas. Fig. 6 presents this network, constructed using VOSviewer.[10] In the visualization, each node represents a keyword, with its size indicating frequency and the edges denoting co-occurrence strength within the same publication. Color-coded clusters group semantically related terms, offering a data-driven overview of the current research agenda and its potential evolution.

The network of co-occurring keywords reveals several distinct clusters that correspond to key themes within the field of sustainability reporting and NLP. The most notable clusters are highlighted by different colors, and their connections provide insight into the relationships between different areas of research. One prominent cluster (colored red) revolves around terms such as "*automated methods*", "*computational linguistics*", "*corporate policies*", and "*knowledge graphs*". This suggests a strong focus on the development and application of advanced computational techniques, including NLP and automated systems, to analyze corporate sustainability reports and extract relevant information. The inclusion of "*knowledge graphs*" points to efforts in structuring and organizing corporate data for better interpretability and decision-making, which are important in automating sustainability report assessments. In the green cluster, terms like "*sustainability*", "*sustainability reporting*", and "*artificial intelligence*" are closely connected. This reflects the growing trend of using AI technologies to automate and enhance the analysis of sustainability reports and CSR efforts. The proximity of terms such as "*artificial neural networks*" and "*machine learning*" in this cluster suggests that much of the work in this area focuses on ML models to process and classify unstructured data from sustainability reports. The blue cluster, which includes terms such as "*general publics*", "*context-aware recommender systems*", and "*stakeholder*", indicates a focus on understanding public and stakeholder perceptions, potentially through recommendation systems designed to align corporate sustainability efforts with stakeholder expectations. The connection to "*sustainability reports*" suggests that these systems might be used to better communicate or tailor sustainability efforts to meet external demands. A distinct orange cluster includes terms such as "*multi-labels*", "*sentiment*

[10] https://www.vosviewer.com/

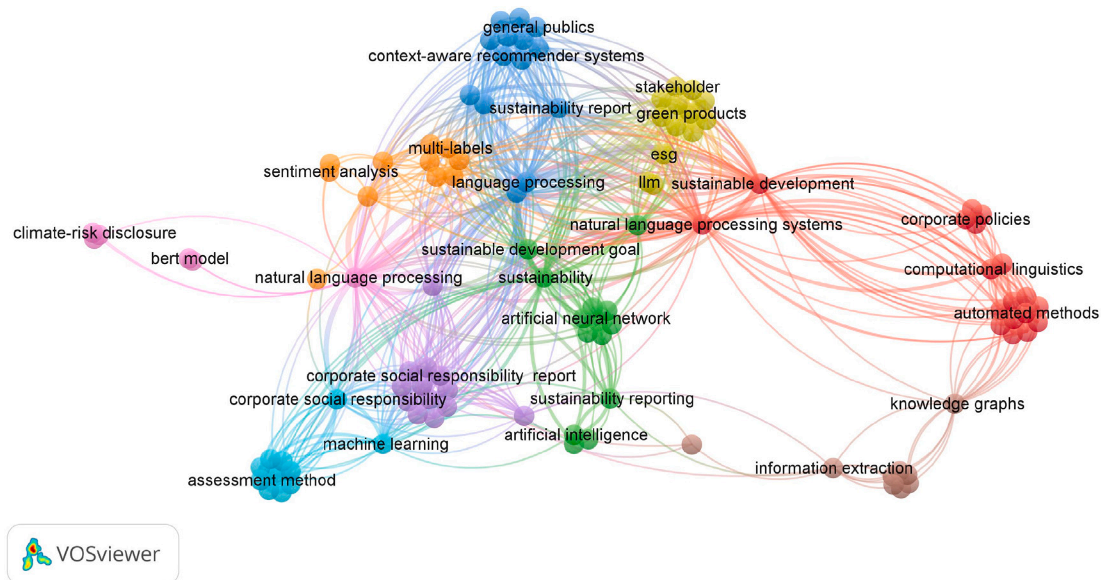

Fig. 6. Network visualization of co-occurring keywords generated using VOSviewer. Each node represents a keyword, with edges indicating co-occurrences within the same paper. Node size reflects the frequency of keyword occurrences, and edge thickness indicates the strength of co-occurrence. Clusters of closely related keywords are distinguished by color.

analysis", and "*natural language processing*", indicating a technical focus on NLP techniques used to classify and analyze sustainability-related text. Sentiment analysis, in particular, is likely used to assess the tone or intent behind sustainability claims, helping to detect greenwashing or misleading statements. Another important connection can be seen in the pink cluster, where terms like "*climate-risk disclosure*" and "*bert model*" co-occur. This points to the use of specific language models, such as BERT, to analyze climate-risk disclosures, reflecting efforts to apply NLP to highly specialized areas of sustainability reporting where precise and reliable information is critical. Finally, the yellow cluster, featuring keywords such as "*ESG*", "*green products*", and "*sustainable development*", highlights research that focuses on ESG factors and the promotion of green products. This cluster is connected to the broader theme of sustainable development, indicating that these studies aim to link corporate sustainability disclosure with global sustainability goals. Overall, the close proximity of these clusters indicates a high level of interdisciplinary research, where advancements in NLP are being applied to diverse aspects of sustainability reporting, from stakeholder engagement to the detection of greenwashing. The co-occurrence of terms suggests that research efforts are converging towards more automated, scalable, and reliable methods for analyzing and improving sustainability practices in line with global goals. The trend analysis of the keywords from the final articles reveals several important insights into the evolving research landscape in sustainability reporting and the application of language technologies. A key observation is the increasing relevance of "*NLP*" for tasks such as "*text processing*" and "*text classification*". NLP techniques, particularly when combined with "*LLMs*", have become critical tools for extracting information and insights from unstructured sustainability reports. This reflects the growing reliance on automated methods to handle the complex and extensive data typically found in sustainability disclosures. At the same time, "*ML*" continues to play a significant role in sustainability analysis, where it is used to detect green claims, improve predictive modeling, and categorize sustainability efforts according to global frameworks such as the "*SDG*", the "*GRI*", and "*ESG*" criteria. The prevalence of terms like "*BERT*" and "*LLMs*" suggests that state-of-the-art models are increasingly being deployed to automate the classification of sustainability reports, aligning corporate claims with these established frameworks. The keyword "*greenwashing*" also appears with notable frequency, underscoring the importance of identifying and mitigating misleading or exaggerated sustainability claims. This suggests that detecting greenwashing has become a key research challenge in the field, where LLMs are applied to assess the validity of corporate sustainability disclosures. The presence of terms like "*sustainability disclosure*" and "*SDG*" in the trend analysis highlights that aligning corporate actions with globally recognized standards remains a fundamental focus of AI applications in this area. Furthermore, as "*AI*" and "*ML*" models become more sophisticated, their role in interpreting, categorizing, and validating these disclosures is expected to expand further. In summary, the trend analysis reflects a shift towards more advanced AI techniques for automating and scaling sustainability reporting, particularly in tasks related to "*greenwashing detection*" and alignment with frameworks like SDGs and GRI. The rising prominence of these topics suggests a growing need for automated solutions that can handle the vast amounts of unstructured data found in sustainability reports while ensuring the accuracy and transparency of corporate sustainability disclosures.

Response to the research Q2. Some of the major challenges and future directions identified in the field, supported by the scientometric analysis (Fig. 6) and the keyword trend analysis, are:

- **Text Extraction and Preprocessing Issues:** One of the significant challenges in sustainability reporting is the text extraction and preprocessing stages. Extracting meaningful text from sustainability reports is often complicated by poor syntactic structures, particularly when dealing with infographics and tables. These reports frequently contain data presented in formats that are not easily readable by standard NLP methods, such as images or graphs, which exacerbates the difficulty of extraction. Preprocessing tasks like stop-word removal and the application of Optical Character Recognition (OCR) are applied but these often fall short in extracting all the relevant information with suitable accuracy. To address these limitations, there is growing interest in multimodal models, such as in [58], to allow direct text extraction from source documents, eliminating the need for OCR. Multimodal approaches enable models to simultaneously process text, images, and other visual data, thereby improving the extraction and understanding of complex formats present in sustainability reports. This leads to more comprehensive and reliable text processing without the traditional reliance on text-based OCR techniques.

- **Data Quantity and Quality Concerns:** A key challenge in training LLMs for sustainability reporting lies in the limited availability of labeled data, which is often expensive and time-consuming to produce. Semi-supervised learning (SSL) offers an effective solution by leveraging large amounts of unlabeled data to reduce the dependency on labeled datasets, thus minimizing costs. The idea behind SSL is to use unlabeled documents – such as sustainability reports that have not been annotated with specific labels – and employ methods as proposed in [59] and [60].
- **Annotation and Domain-Specific Challenges:** Annotating sustainability reports in alignment with established frameworks like GRI, SDG, and ESG presents a major challenge due to the complexity of these frameworks, as well as the contextual and domain-specific knowledge required for accurate annotation. This process is not only time-consuming and costly but also prone to inconsistencies if done manually, as it requires deep expertise to correctly interpret and map text segments to the appropriate categories within each framework. To address this challenge, there is significant potential in creating shared, cross-domain annotated datasets that integrate aspects of all three frameworks (GRI, SDG, ESG). By combining these annotations, researchers and practitioners can exploit the natural overlaps and interconnections between the frameworks. For example, a company's ESG performance can often be linked to specific SDG targets or GRI reporting standards, creating opportunities for cross-referencing and harmonizing different sustainability dimensions. Such an approach would improve the consistency of annotations, making them more reusable across tasks, models, and studies. Furthermore, developing a unified annotation schema across the three frameworks would allow for more efficient cross-domain training of models, leveraging relationships between economic, environmental, and social dimensions.
- **Model and Methodological Limitations for Greenwashing Detection:** Current LLMs face limitations in detecting nuanced language used in greenwashing, as the training data may not be diverse or extensive enough to cover the full range of deceptive strategies. For example, models may struggle with context loss when analyzing sustainability reports paragraph-by-paragraph, missing the broader patterns of misleading claims. To overcome this challenge, models need more sophisticated contextual awareness, as well as datasets that encompass diverse examples of greenwashing across various industries and regions. Fine-tuning pre-trained models on domain-specific corpora, or integrating external data sources such as news articles and social media, could enhance the detection of greenwashing claims. Additionally, more work is needed in multi-modal learning where textual and non-textual information (e.g., images and graphs) are analyzed together, which would help capture misleading claims presented in different forms.
- **Scalability, Generalization, and Benchmarking of Models**: Current LLMs applied to sustainability reporting and greenwashing detection often lack the ability to generalize across diverse industries, regions, and languages, limiting their broader applicability. A major issue is the absence of standardized benchmarks and shared datasets that could allow for a meaningful comparison between models. Most studies focus on specific sectors or geographic contexts, making it difficult to assess how these models perform in varying conditions. To address this, future research should prioritize the development of open, shared datasets that encompass multiple industries and regulatory frameworks. These datasets would enable consistent benchmarking of different models, such as BERT, RoBERTa, GPT, and others, across tasks like greenwashing detection and sustainability report classification.
- **Regulatory and Ethical Concerns:** As language technologies become more central to sustainability reporting and greenwashing detection, ethical considerations around transparency, accountability, and fairness are increasingly important. Models must ensure fairness in decision-making processes, avoid biases related to specific regions or industries, and adhere to global sustainability standards. In addition, regulators and policymakers need to establish frameworks for how these LLMs-driven approaches can be integrated into formal sustainability auditing processes, ensuring their outputs are both reliable and aligned with international standards like the GRI, SDG, and ESG. Collaborative efforts between AI developers, regulators, and industry stakeholders will be essential to address these challenges.
- **Integration of External Data Sources:** Many sustainability reports are self-published by companies, making them prone to biases or incomplete data that can obscure true sustainability performance. Integrating external data sources such as media reports, regulatory filings, and third-party audits can provide a more balanced and accurate analysis. NLP and LLM models can be enhanced by incorporating external sources to cross-reference corporate claims, thereby improving the reliability of greenwashing detection and sustainability assessments. For example, external data can offer valuable context for validating or questioning the claims made in sustainability reports, creating a more robust evaluation framework.

Future directions. Future research directions after reviewing the articles are:

- **Improved Annotation and Parsing:** Enhancing the annotation process is crucial to better align text from sustainability reports with the relevant frameworks (e.g., GRI, SDG, ESG). This can be achieved by advancing parsing techniques that can handle not just simple text but also information embedded in infographics, tables, and charts. A promising direction involves integrating a core Semantic Role Labeling (SRL) stage [61], which produces more detailed information about the roles played by the different entities in a sentence. By assigning semantic roles, such as "agent", "action", and "object", to various parts of the text, the extracted structured information becomes richer and more actionable. Future work could also explore extending this annotation to include multi-modal data from documents that contain both text and visual elements, improving the overall extraction and understanding process.
- **User Feedback and Data Augmentation:** One important future direction is the integration of user feedback into model improvement cycles. By continuously gathering feedback from users (e.g., sustainability analysts or auditors) who interact with AI systems, the performance of LLMs can be incrementally improved. Furthermore, this feedback can be used to augment existing datasets with additional annotated data. Techniques like crowdsourcing could be used to annotate documents in a scalable way, or active learning methods could be applied, where the model itself helps identify the most informative data points to annotate next. This user-centric, iterative approach ensures that models evolve in response to real-world demands and become more accurate over time. Additionally, future studies should collect data regarding the real-world impact of language technologies on sustainability assessments to inform model development and evaluation.
- **Time-Series Studies:** While current research often focuses on static analyses, there is a growing need for time-series studies to understand the longitudinal behavior of organizations. By tracking the sustainability performance of companies over time, researchers can identify trends, shifts in corporate behavior, and the long-term impacts of sustainability strategies. Such studies would also enable the detection of changes in greenwashing practices or improvements in compliance with sustainability goals.

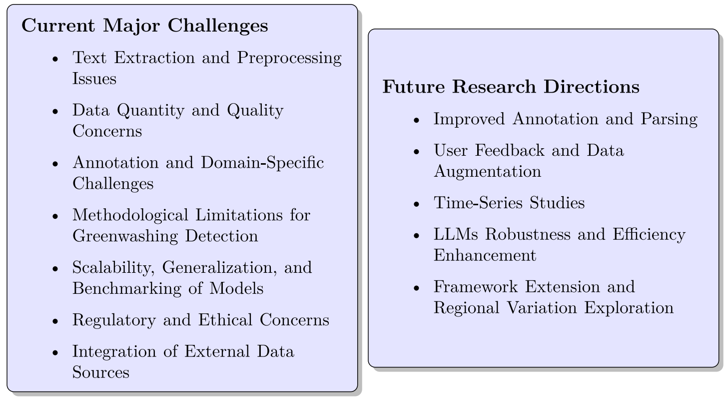

Fig. 7. Overview of current challenges and future research directions in sustainability reporting with LLMs.

Advanced techniques, such as dynamic LLMs that evolve with time-series data, could be developed to analyze historical reports and predict future sustainability actions or risks.

- **LLMs Robustness and Efficiency Enhancement:** As LLMs become more integrated into sustainability reporting analysis, it is essential to continuously improve their robustness and efficiency. Research should focus on developing models that are more resilient to noisy, incomplete, or inconsistent data, such as poorly formatted reports or documents with missing sections. Additionally, enhancing the efficiency of LLMs is critical for scaling these models to analyze vast amounts of data in real time, especially as more organizations are required to publish sustainability reports. Techniques like model pruning, quantization, and knowledge distillation can be explored to reduce the computational footprint of LLMs without sacrificing performance. Comparative studies between different LLM architectures (e.g., GPT-4, BERT variants) should also be conducted to identify the best models for specific sustainability tasks, such as greenwashing detection or sentiment analysis.
- **Framework Extension and Regional Variation Exploration:** As sustainability reporting becomes more globalized, it is important to extend current frameworks (GRI, SDG, ESG) to accommodate regional variations and multi-language capabilities. Many existing models are limited by geographic biases, often being trained on datasets from specific countries or industries. Future research should focus on broadening the scope of these models to handle regional differences in ESG sentiment and regulatory requirements. This will involve extending sustainability frameworks to include reports in multiple languages and from diverse regions. Developing models that can automatically detect and adjust for regional variations in sustainability reporting, such as differing environmental regulations or cultural perspectives on social responsibility, will be a key step towards making these systems more globally applicable.

Fig. 7 provides a side-by-side overview of the current challenges and future research directions in applying LLMs to sustainability reporting, highlighting the shift from foundational technical and data issues towards enhancing model robustness, user interaction, and contextual adaptability.

6. Conclusions

Sustainability reporting is central to promoting transparency, accountability, and alignment with global frameworks such as the SDGs, GRI, and ESG. Yet, persistent challenges, ranging from the complexity of unstructured data to the risks of greenwashing, continue to hinder the reliability and comparability of disclosures. This study presented a systematic literature review of 21 relevant articles, analyzing how language technologies, particularly LLMs, are being used to address two primary challenges: extracting structured insights from sustainability reports and detecting misleading claims. The review highlights the increasing adoption of LLMs for tasks such as classification, information extraction, and claim verification. However, several critical limitations remain, including difficulties in preprocessing complex report formats, the scarcity of high-quality annotated data, and the limited ability of existing models to detect nuanced forms of greenwashing. These findings point to the need for more sophisticated multimodal models, improved annotation workflows, and the development of shared, cross-framework datasets. Future research should prioritize enhancing the robustness and adaptability of LLMs, integrating semantic role labeling, incorporating user feedback, and addressing regional and linguistic diversity in reporting practices. By advancing these methodological fronts, language technologies can play a transformative role in making sustainability reporting more scalable, interpretable, and trustworthy, ultimately supporting more informed, accountable, and globally aligned business decision-making.

This review identifies priority avenues for strengthening the evidence base. First, the limited material for Challenge 2 indicates the need for more systematic, model-agnostic studies on greenwashing detection. Second, while findings tied to specific LLM families or versions are time-bound, this reflects a broader pattern of rapid, cumulative progress that is expanding methodological options. Third, constraints on reproducibility, stemming from proprietary systems and scarce open ESG corpora, underscore the value of open datasets, shared protocols, and collaborative infrastructures. Progress on these priorities can enhance robustness, comparability, and practical utility in sustainability reporting.

CRediT authorship contribution statement

Seyed Alireza Mousavian Anaraki: Writing – review & editing, Writing – original draft, Conceptualization. **Danilo Croce:** Writing – review & editing, Writing – original draft, Conceptualization. **Roberto Basili:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

We have nothing to declare.

Acknowledgments

We have no acknowledgments to declare.

Appendix A. Overview of reporting frameworks: GRI, SDGs, and ESG

This appendix provides an extended overview of the key sustainability frameworks referenced in the main text, namely, GRI, SDG, and ESG. The focus is on their structure, use in reporting practices, and relevance to the literature analyzed in this review.

The GRI is recognized for creating and promoting universally applicable standard for sustainability reporting. It comprises Universal Standards (Codes 101–103) and three sets of topic-specific standards: Economic (Codes 201–207), Environment (Codes 301–308), and Social (Codes 401–418). Universal Standards lay the foundation for creating sustainability reports by offering essential standards. They encompass information on an organization's operational context, strategy, governance, and stakeholder engagement. For instance, GRI 102 provides General Disclosures, detailing aspects such as "*Organization's profile, Strategic analysis, and Governance structure*". Topic-specific standards, on the other hand, delve into particular areas of sustainability, offering comprehensive guidance on reporting these aspects. For example in Environment, GRI 303 pertains to Water and Effluents, advising organizations on reporting their water usage and effluent management practices.

These standards often contain detailed sub-codes to facilitate more accurate and specific disclosures. As an example, the Social code 401-Employment includes three sub-codes: 401-1 for "*New employee hires and employee turnover*", 401-2 for "*Benefits provided to full-time employees that are not offered to temporary or part-time employees*", and 401-3 for "*Parental leave*". Each code outlines specific reporting requirements, recommendations, and guidance [4].

The UN SDGs encompass 17 Global Goals (detailed in Table A.3) and 169 specific targets designed to tackle major development challenges and advance global sustainability [13]. These targets provide a detailed description of what is intended to be achieved with each goal. Each of them is accompanied by indicators that serve to measure and monitor progress towards the achievement of each target. This structure allows for a clear and measurable picture of the actions to be taken and the progress made [3]. For instance, the first goal (SDG 1, also reported in Table A.3) pertains to "*Ending poverty in all its forms everywhere*", encouraging companies to report on their initiatives aimed at alleviating poverty and improving the living conditions of the most vulnerable communities. The SDGs, though individually distinct, are closely interconnected and often influence each other. This interdependence emphasizes the need for an integrated approach to sustainable development. Progress in one goal can have either positive or negative impacts on others, highlighting the importance of considering the 2030 Agenda in its entirety, rather than viewing the goals as separate and independent entities.

For instance, the interconnection between 1-Poverty, 3-Health, and 4-Education demonstrates the intertwined nature of the SDGs. The fight against 1-Poverty is closely related to 3-Health and 4-Education. People living in poverty often face restricted access to healthcare and education, which can perpetuate a cycle of disadvantage. Conversely, education equips individuals with the skills and knowledge necessary to enhance their health and economic conditions, thereby breaking the cycle of poverty [62].

The performance of firms aligned with ESG factors provides investors with greater confidence in the authenticity of their sustainability efforts. Various ESG standards, such as SASB, TCFD, and GRI, guide corporate decision-making, each focusing on different aspects of sustainability [7].

The SASB, for instance, is a non-profit organization that develops and disseminates sustainability standards specifically for companies to disclose information to investors. The SASB's standard predominantly addresses the U.S. context and prioritizes data that is most relevant to investors. On the other hand, the TCFD, established by the Financial Stability Board, provides voluntary recommendations focused exclusively on climate-related financial disclosures. Unlike the TCFD's narrow focus on climate issues, the GRI offers comprehensive coverage of a broader range of sustainability topics.

Appendix B. Typologies and examples of Greenwashing

This appendix expands on the concept of greenwashing as discussed in the main text, focusing on its categorization across ESG factors, as well as its manifestations at both the firm and product levels. It also presents a classification of green claims based on their type and degree of deceptiveness. Environmental greenwashing is the most common form and occurs when a company provides false or misleading information about its environmental performance or practices, such as energy efficiency, resource use, or waste management. Social greenwashing can occur when a company exaggerates or falsifies its social credentials, such as working conditions, fair hiring practices, or community initiatives. Governance greenwashing can occur when a company provides false or misleading information about its governance practices, such as transparency, business ethics, or risk management [47].

de Freitas Netto et al. [10] classify greenwashing into two primary forms: Claim greenwashing and Executional greenwashing. Both of which pertain to firm-level and product/service-level activities. Their findings indicate that most research has focused on product/service-level claim greenwashing, where textual assertions explicitly or implicitly suggest ecological benefits to create misleading environmental claims. These "green claims" are divided into two types: (*i*) Claim type and (*ii*) Claim deceptiveness [63]. Each of these types is further categorized into five categories, as explained in Table B.4.

The Volkswagen scandal is a key example of greenwashing, specifically illustrating Process Orientation and False/Outright Lie in the context of deceptive green claims. *Volkswagen installed deception devices in 590,000 diesel cars to manipulate Environmental Protection Agency (EPA) emissions tests, falsely representing the environmental performance of its vehicles. This fraudulent action led Volkswagen to plead guilty, pay a 2.8 billion dollar fine, and suffer 7 billion Euro in profit losses, a 25% drop in share value, and reduced investments in 2015* [28].

Appendix C. Literature search process and query table

Table C.5 reports the exact logical queries used to search the WoS and Scopus databases. These queries were designed to retrieve studies that explore the application of NLP and LLMs in the context of sustainability reporting, including frameworks such as ESG, GRI, and SDG. The Table provides the detailed logic applied in each database, along with the number of records retrieved. The queries combine terms related to various types of sustainability reports (e.g., "sustainability report", "ESG disclosure", "CSR report") with NLP/LLM-related keywords.

Appendix D. Overview of excluded papers and rationale

This appendix provides an overview of the main topical themes among the excluded papers, along with representative examples. These cases help clarify the scope of the review and illustrate why certain studies were excluded.

In the first phase of the filtering process, papers were excluded based on title relevance, as many focused on broad applications of AI techniques to enhance sustainability in various specialized domains. It's noteworthy that several of these works, while employing models like BERT and other advanced NLP techniques, did not center on the analysis of sustainability reports. They focused, instead, on the development of sustainable practices within their respective fields.

Here are some of the main topics of such excluded papers, along with examples of the most relevant works:

Table A.3
Abbreviation and descriptions of SDGs.
Source: [3].

No.	Abbreviation	Description
1	Poverty	*End poverty in all its forms everywhere*
2	Hunger	*End hunger, achieve food security and improved nutrition, and promote sustainable agriculture*
3	Health	*Ensure healthy lives and promote well-being for all at all ages*
4	Education	*Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all*
5	Gender	*Achieve gender equality and empower all women and girls*
6	Sanitation	*Ensure availability and sustainable management of water and sanitation for all*
7	Energy	*Ensure access to affordable, reliable, sustainable, and modern energy for all*
8	Economy	*Promote sustained, inclusive, and sustainable economic growth, full and productive employment, and decent work for all*
9	Industry	*Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation*
10	Inequality	*Reduce inequality within and among countries*
11	Settlements	*Make cities and human settlements inclusive, safe, resilient, and sustainable*
12	Consumption	*Ensure sustainable consumption and production patterns*
13	Climate	*Take urgent action to combat climate change and its impacts*
14	Aquatic	*Conserve and sustainably use the oceans, seas, and marine resources for sustainable development*
15	Terrestrial	*Protect, restore, and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, halt and reverse land degradation, and halt biodiversity loss*
16	Peace	*Promote peaceful and inclusive societies for sustainable development, provide access to justice for all, and build effective, accountable, and inclusive institutions at all levels*
17	Partnerships	*Strengthen the means of implementation and revitalize the global partnership for sustainable development*

Table B.4
Green claims types.
Source: [10].

Green claims types	Categories	Explanation
Claim Type	Product Orientation	*Concentrating on the ecological characteristics of a product.*
	Process Orientation	*Highlighting the environmental performance of a manufacturing process, technique, or disposal method.*
	Image Orientation	*Aiming to enhance the environmentally-friendly reputation of a company by associating it with environmental causes or activities with high public support.*
	Environmental Fact	*Presenting an ostensibly factual statement from a company regarding the environment or its status.*
	Combination	*Incorporating two or more of the aforementioned categories.*
Claim Deceptiveness	Vague/Ambiguous	*Claims that are overly vague, ambiguous, broad, or lacking a clear definition.*
	Omission	*Claims that omit essential information to assess their validity.*
	False/Outright Lie	*Claims that are inaccurate or fabricated.*
	Combination	*Claims that fall into two or more of the aforementioned deceptive categories.*
	Acceptable	*Claims that do not contain deceptive features.*

- High-Level Introduction to AI in Sustainability. Chen et al. [64] reviewed significant trends in ML aimed at addressing sustainability challenges in AI, particularly focusing on environmental and social aspects. Vaiyapuri et al. [65] conducted an AI-based sentiment analysis on Twitter data related to the COVID-19 pandemic, demonstrating the use of sustainable AI methods (such as BERT) but not over Sustainability Reports.
- AI in Healthcare and Sustainable Practices. Similarly, AI has been applied in various studies within the healthcare domain to develop sustainable practices. For instance, Gaudet-Blavignac et al. [66] focused on optimizing health data usage through semantic AI, aiming to enhance sustainable health practices. Fukuoka et al. [67] employed NLP to analyze motivational profiles, contributing to the maintenance of physical activity and the development of sustainable healthcare interventions.
- AI for Enhancing Sustainability in Specialized Domains. While the previous group of studies focuses on applications within the healthcare sector, there are also works dedicated to enhancing sustainability in other specific domains. For example, Khairuddin et al. [68] utilized AI to assess sustainability and safety in occupational health. Similarly, Usigbe et al. [69] applied AI-based technologies to improve resilience and sustainability in agricultural production systems, aiming to optimize processes, enhance food security, and reduce environmental emissions. These studies, however, do not directly relate to sustainability reporting but instead focus on applying AI to address sustainability challenges

Table C.5
Search queries and number of documents retrieved from WoS and Scopus.

Search domain	Logical query	No. of docs
WoS	((TS = (*sustainability report*) AND TS = (*Natural language processing*)) OR (TS = (*sustainability report*) AND TS = (*large language model*)) OR (TS = (*ESG report*) AND TS = (*Natural language processing*)) OR (TS = (*ESG report*) AND TS = (*large language model*)) OR (TS = (*CSR report*) AND TS = (*Natural language processing*)) OR (TS = (*CSR report*) AND TS = (*large language model*)) OR (TS = (*GRI report*) AND TS = (*Natural language processing*)) OR (TS = (*GRI report*) AND TS = (*large language model*)) OR (TS = (*SDG report*) AND TS = (*Natural language processing*)) OR (TS = (*SDG report*) AND TS = (*large language model*)) OR (TS = (*ESG disclosure*) AND TS = (*Natural language processing*)) OR (TS = (*ESG disclosure*) AND TS = (*large language model*)) OR (TS = (*sustainability disclosure*) AND TS = (*Natural language processing*)) OR (TS = (*sustainability disclosure*) AND TS = (*large language model*)) OR (TS = (*Greenwashing*) AND TS = (*Natural language processing*)) OR (TS = (*Greenwashing*) AND TS = (*large language model*)))	68
Scopus	((TITLE-ABS-KEY (*sustainability AND report*) AND TITLE-ABS-KEY (*natural AND language AND processing*)) OR (TITLE-ABS-KEY (*sustainability AND report*) AND TITLE-ABS-KEY (*large AND language AND model*)) OR (TITLE-ABS-KEY (*esg AND report*) AND TITLE-ABS-KEY (*natural AND language AND processing*)) OR (TITLE-ABS-KEY (*esg AND report*) AND TITLE-ABS-KEY (*large AND language AND model*)) OR (TITLE-ABS-KEY (*csr AND report*) AND TITLE-ABS-KEY (*natural AND language AND processing*)) OR (TITLE-ABS-KEY (*csr AND report*) AND TITLE-ABS-KEY (*large AND language AND model*)) OR (TITLE-ABS-KEY (*gri AND report*) AND TITLE-ABS-KEY (*natural AND language AND processing*)) OR (TITLE-ABS-KEY (*gri AND report*) AND TITLE-ABS-KEY (*large AND language AND model*)) OR (TITLE-ABS-KEY (*sdg AND report*) AND TITLE-ABS-KEY (*natural AND language AND processing*)) OR (TITLE-ABS-KEY (*sdg AND report*) AND TITLE-ABS-KEY (*large AND language AND model*)) OR (TITLE-ABS-KEY (*esg AND disclosure*) AND TITLE-ABS-KEY (*natural AND language AND processing*)) OR (TITLE-ABS-KEY (*esg AND disclosure*) AND TITLE-ABS-KEY (*large AND language AND model*)) OR (TITLE-ABS-KEY (*sustainability AND disclosure*) AND TITLE-ABS-KEY (*natural AND language AND processing*)) OR (TITLE-ABS-KEY (*sustainability AND disclosure*) AND TITLE-ABS-KEY (*large AND language AND model*)) OR (TITLE-ABS-KEY (*greenwashing*) AND TITLE-ABS-KEY (*natural AND language AND processing*)) OR (TITLE-ABS-KEY (*greenwashing*) AND TITLE-ABS-KEY (*large AND language AND model*)))	89
Unique Papers		136
Other sources		4
Total		140

within specific fields. For similar reasons, the conceptual works by [20,21] were not included in the final selection. Boedijanto and Delina [21] concentrates specifically on greenwashing detection within the energy sector, limiting its applicability to broader sustainability reporting practices. De Villiers et al. [20] focuses on conceptual implications of generative AI in areas like sustainability accounting and assurance, without offering empirical analysis.

In the second phase of the filtering process, the abstracts of the papers were examined more closely. Unlike the first group, which focused on applying AI to develop sustainable practices across various domains, this second group includes works that specifically engage with sustainability documents. These studies explore the limitations and opportunities inherent in these reports but do not apply advanced NLP or LLM techniques as examined in this review. Instead, they focus on assessing and analyzing aspects such as readability, complexity, and integration into management practices, using more traditional analytical methods. Examples include:

- Measuring CSR Readability by AI. Smeuninx et al. [70] investigated the readability of sustainability reports, finding them more challenging to read than financial reports, despite being targeted at a broader audience. This work did not utilize NLP methods for extracting or analyzing data but rather focused on readability metrics.
- Measuring Textual Complexity of CSR by AI. Almendros et al. [71] analyzed the simplicity and redundancy of CSR reports compared to standard language, highlighting that these reports are often designed more for shareholders than the general public. Their approach was centered around assessing textual complexity without using advanced NLP techniques.
- Integration of Sustainability Reporting into Management Practices. Adams and Frost [72] explored the development and application of KPI for sustainability, emphasizing their role in decision-making, planning, and performance management. Similarly Hamza and Jarboui [73] investigated the impact of board composition on ESG performance and disclosure quality, particularly focusing on impression management. These studies primarily addressed the integration of sustainability into corporate governance rather than employing NLP methods.

In the final phase of the filtering process, we conducted an in-depth review of the full content of each paper. This was necessary because the abstracts did not always provide enough detail to fully understand the methodologies used. Through this comprehensive analysis, it became clear that some papers, while initially appearing relevant, employed techniques such as topic modeling, visual content analysis, or basic text mining. However, they did not utilize advanced NLP methods, which are the central focus of this review. As a result, we excluded papers that were oriented around:

- Topic Modeling and CSR Theme Development in Sustainability. While papers like [74] explored topics related to sustainability, distributing them across environmental, social, economic, and general factors, and [75] focused on developing CSR themes using text mining and topic modeling [76], they did not employ the advanced NLP or LLM techniques pertinent to this review. Their methods were primarily oriented around thematic development rather than extracting or analyzing complex textual data using modern AI techniques.
- Investigating the Impact of Visual Content on CSR using AI. Nakao et al. [77] examined corporate reporting strategies through the lens of visual content analysis. Although this paper utilized AI, it was more aligned with visual data analysis rather than the techniques central to our research.
- AI-Enhanced Sustainability Indicators. Papers such as [78,79] utilized text mining to derive sustainability indicators, analyzing environmental complaints and unstructured digital news articles,

respectively. However, their focus was on using AI to track specific metrics and indicators rather than leveraging NLP and LLMs to provide a comprehensive analysis of sustainability reporting practices.

Excluded Review Articles. During the screening phase, several review papers were excluded from the final analysis as they did not meet the inclusion criteria, particularly the explicit focus on the application of NLP or LLM techniques to sustainability reporting. While these papers addressed important sustainability-related topics, they were not sufficiently aligned with the technological scope of this review. Below are examples of excluded review papers and their thematic focus:

- Geissdoerfer et al. [80] examined the conceptual overlap between Circular Economy and sustainability using bibliometric and snowballing techniques.
- Székely and Vom Brocke [74] explored the role of NLP as an investigative tool in sustainability studies but lacked direct application to sustainability reporting.
- Velte [81] focused on Automated Text Analyses (ATAs) for integrated reporting but without employing LLM-based techniques.
- Lim [19] reviewed the link between ESG, AI, and finance, identifying key applications of AI in ESG disclosure but not emphasizing NLP or sustainability report analysis.
- Castelo Branco [82] offered a meta-review of CSR-related accounting research and emphasized the need for critical engagement with the literature.
- Cunha and Duncan [83] conducted a bibliometric analysis on integrated reporting and decision-making but did not discuss NLP methods or textual processing techniques.

These papers, while informative, were excluded as they did not directly analyze sustainability reports using advanced NLP or LLM approaches. They remain valuable references for broader sustainability and AI research.

Appendix E. Critical appraisal checklists used for quality assessment

To ensure methodological rigor and transparency, all articles included in this review were systematically evaluated using established critical appraisal checklists, adapted from the CASP. The checklists were tailored to the specific type of publication (research article or review article) and guided the assessment of study design, data quality, and reporting standards. For the 20 considered Research papers, the following checklist was used to evaluate their quality:

1. **Research Objectives:** *Are the goals of the research explicitly stated?*
2. **Research Methodology:** *Is research methodology appropriate?*
3. **Research Design:** *Is the research design appropriate to address the research aims?*
4. **Sampling Strategy:** *Is the sampling strategy appropriate to address the research aims?*
5. **Data Collection:** *Is the data collected in a manner that addresses the research issue?*
6. **Researcher–Participant Relationship:** *Has the researcher-participant relationship and potential biases been identified?*
7. **Ethical Considerations:** *Are ethical considerations taken into account during the research?*
8. **Accuracy of Data Analysis:** *Is the data analysis sufficiently meticulous?*
9. **Clarity of findings:** *Are the findings of the research clearly stated?*
10. **Overall Research Value:** *Is the research valuable?*

For the Review paper, the following checklist was used to evaluate its quality

1. **Research Question:** *Did the review tackle a specific, well-defined question?*
2. **Paper Selection:** *Did the authors target the correct types of studies for inclusion?*
3. **Relevant Studies:** *Were all significant and relevant studies included in the review?*
4. **Study Quality:** *Did the authors adequately evaluate the quality of the studies included?*
5. **Combining Results:** *If the results of the review have been combined, was this approach justified?*
6. **Clarity of Results:** *Are the results of the review clear?*
7. **Precision of Results:** *Are the results accurate?*
8. **Applicability:** *Are the findings applicable to the local population?*
9. **Important Outcomes:** *Were all crucial outcomes considered?*
10. **Benefits and Costs:** *Are the benefits worth the harms and costs?*

For each criterion, responses were recorded as "*Yes*", "*No*", or "*Can't tell*". Answering "*Yes*" indicates that the article fully meets the criterion; "*No*" highlights a clear limitation or omission; while "*Can't tell*" is used when the information provided is insufficient or unclear to make a judgment. This structured appraisal not only ensures consistency and transparency in the evaluation process, but also strengthens the reliability and comparability of the review's findings.

Data availability

No data was used for the research described in the article.

References

[1] M. Angin, B. Taşdemir, C.A. Yılmaz, G. Demiralp, M. Atay, P. Angin, G. Dikmener, A roberta approach for automated processing of sustainability reports, Sustainability 14 (16139) (2022) http://dx.doi.org/10.3390/su142316139.
[2] W. Moodaley, A. Telukdarie, Greenwashing, sustainability reporting, and artificial intelligence: A systematic literature review, Sustainability 15 (2023b) 1481, http://dx.doi.org/10.3390/su15021481.
[3] U. Nations, Transforming our world: The 2030 agenda for sustainable development, vol. 1, Department of Economic and Social Affairs, New York: United Nations, 2015, p. 41.
[4] I. Nechaev, D.S. Hain, Social impacts reflected in csr reports: Method of extraction and link to firms' innovation capacity, J. Clean. Prod. 429 (2023) 139256, http://dx.doi.org/10.1016/j.jclepro.2023.139256.
[5] S. Ghosh, S.K. Naskar, Ranking environment, social and governance related concepts and assessing sustainability aspect of financial texts, in: Proc. 4th Workshop Fintech Nat. Lang. Process., FinNLP, 2022, pp. 243–249, http://dx.doi.org/10.18653/v1/2022.finnlp-1.33.
[6] V. Bogdan, L. Rus, D.S. Gherai, A.G. Florea, N.G. Bugnar, A streamline sustainable business performance reporting model by an integrated finesg approach, Sustainability 15 (16860) (2023) http://dx.doi.org/10.3390/su152416860.
[7] Y. Zou, M. Shi, Z. Chen, Z. Deng, Z. Lei, Z. Zeng, S. Yang, H. Tong, L. Xiao, W. Zhou, Esgreveal: An llm-based approach for extracting structured data from esg reports, 2023, http://dx.doi.org/10.48550/arXiv.2312.17264, arXiv preprint arXiv:2312.17264.
[8] M. Bronzini, C. Nicolini, B. Lepri, A. Passerini, J. Staiano, Glitter or gold? deriving structured insights from sustainability reports via large language models, EPJ Data Sci. 13 (41) (2024) http://dx.doi.org/10.48550/arXiv.2310.05628.
[9] S. Schmidt, J. Kinne, S. Lautenbach, T. Blaschke, D. Lenz, B. Resch, 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. 835 (2022) 155512, http://dx.doi.org/10.1016/j.scitotenv.2022.155512.
[10] S.V. de Freitas Netto, M.F.F. Sobral, A.R.B. Ribeiro, G.R.L. Soares, Concepts and forms of greenwashing: A systematic review, Env. Sci. Eur. 32 (2020) 1–12, http://dx.doi.org/10.1186/s12302-020-0300-3.
[11] M. Polignano, N. Bellantuono, F.P. Lagrasta, S. Caputo, P. Pontrandolfo, G. Semeraro, An nlp approach for the analysis of global reporting initiative indexes from corporate sustainability reports, in: Proc. 1st Comput. Soc. Responsib. Workshop Within 13th Lang. Resour. Eval. Conf., 2022, pp. 1–8.
[12] A. Vinella, M. Capetz, R. Pattichis, C. Chance, R. Ghosh, Leveraging language models to detect greenwashing, 2023, http://dx.doi.org/10.48550/arXiv.2311.01469, arXiv preprint arXiv:2311.01469.

[13] T.B. Smith, R. Vacca, L. Mantegazza, I. Capua, Natural language processing and network analysis provide novel insights on policy and scientific discourse around sustainable development goals, Sci. Rep. 11 (2021) 22427, http://dx.doi.org/10.1038/s41598-021-01801-6.

[14] S. Minaee, T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, J. Gao, Large language models: A survey, 2024, http://dx.doi.org/10.48550/arXiv.2402.06196, arXiv preprint arXiv:2402.06196.

[15] W.X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J.Y. Nie, J.R. Wen, A survey of large language models, 2023, https://arxiv.org/abs/2303.18223, arXiv:2303.18223.

[16] W. Moodaley, A. Telukdarie, A conceptual framework for subdomain specific pre-training of large language models for green claim detection, Eur. J. Sustain. Dev. 12 (2023a) http://dx.doi.org/10.14207/ejsd.2023.v12n4p319, 319–319.

[17] H. Diwan, B.Amarayil. Sreeraman, From financial reporting to esg reporting: a bibliometric analysis of the evolution in corporate sustainability disclosures, Env. Dev. Sustain. 26 (2024) 13769–13805, http://dx.doi.org/10.1007/s10668-023-03249-2.

[18] O.A. Osobajo, A. Oke, A. Lawani, T.S. Omotayo, N. Ndubuka-McCallum, L. Obi, Providing a roadmap for future research agenda: A bibliometric literature review of sustainability performance reporting (spr), Sustainability 14 (8523) (2022) http://dx.doi.org/10.3390/su14148523.

[19] T. Lim, Environmental, social, and governance (esg) and artificial intelligence in finance: State-of-the-art and research takeaways, Artif. Intell. Rev. 57 (2024) 76, http://dx.doi.org/10.1007/s10462-024-10708-3.

[20] C. De Villiers, R. Dimes, M. Molinari, How will ai text generation and processing impact sustainability reporting? critical analysis, a conceptual framework and avenues for future research, Sustain. Account. Manag. Policy J. 15 (2024) 96–118.

[21] F.J.O. Boedijanto, L.L. Delina, Potentials and challenges of artificial intelligence-supported greenwashing detection in the energy sector, Energy Res. Soc. Sci. 115 (2024) 103638.

[22] I. KPMG, The Kpmg Survey of Corporate Responsibility Reporting 2013, KPMG Int. Zurich, Switzerland, 2013.

[23] R.G. Eccles, G. Serafeim, D. Seth, C.C.Y. Ming, The performance frontier: Innovating for a sustainable strategy: Interaction, Harv. Bus. Rev. 91 (2013) 17–18.

[24] A. Kolk, The social responsibility of international business: From ethics and the environment to csr and sustainable development, J. World Bus. 51 (2016) 23–34, http://dx.doi.org/10.1016/j.jwb.2015.08.010.

[25] A. Gupta, A. Chadha, V. Tewari, A natural language processing model on bert and yake technique for keyword extraction on sustainability reports, IEEE Access (2024) http://dx.doi.org/10.1109/ACCESS.2024.3352742.

[26] T.P. Lyon, J.W. Maxwell, Greenwash: Corporate environmental disclosure under threat of audit, J. Econ. Manag. Strat. 20 (2011) 3–41, http://dx.doi.org/10.1111/j.1530-9134.2010.00282.x.

[27] P.R. Milgrom, J. Roberts, Economics, Organization and Management, Prentice-Hall, Englewood Cliffs, NJ, 1992.

[28] T. Pimonenko, Y. Bilan, J. Horák, L. Starchenko, W. Gajda, Green brand of companies and greenwashing under sustainable development goals, Sustainability 12 (1679) (2020) http://dx.doi.org/10.3390/su12041679.

[29] A.Y. Zhang, J.H. Zhang, Renovation in environmental, social and governance (esg) research: the application of machine learning, Asian Rev. Account. 32 (2024) 554–572.

[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Adv. Neural Inf. Process. Syst, vol. 30, 2017.

[31] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, X. Huang, Pre-trained models for natural language processing: A survey, Sci. China Technol. Sci. 63 (2020) 1872–1897, http://dx.doi.org/10.1007/s11431-020-1647-3.

[32] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.

[33] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, 2019.

[34] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving language understanding by generative pre-training, 2018.

[35] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, Language models are unsupervised multitask learners, OpenAI Blog 1 (9) (2019).

[36] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, G. Lample, Llama: Open and efficient foundation language models, 2023, arXiv:2302.13971.

[37] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P.J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, J. Mach. Learn. Res. 21 (2020) 1–67.

[38] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, 2019, arXiv preprint arXiv:1910.13461.

[39] F. Gonzalez, Z. Jin, B. Schölkopf, T. Hope, M. Sachan, R. Mihalcea, Beyond good intentions: Reporting the research landscape of nlp for social good, 2023, http://dx.doi.org/10.48550/arXiv.2305.05471, arXiv preprint arXiv:2305.05471.

[40] C. Jakob, V. Schmitt, S. Mohtaj, S. Möller, Classifying sustainability reports using companies' self-assessments, in: Future of Information and Communication Conference, Springer, 2024, pp. 547–557, http://dx.doi.org/10.1007/978-3-031-53960-2_35.

[41] M. Sandelowski, J. Barroso, Handbook for Synthesizing Qualitative Research, springer publishing company, 2006.

[42] S. Rouhani, S. Rotbei, M. Shamizanjani, Meta-synthesis of big data impacts on information systems development, J. Manag. Anal. 4 (2017) 182–201, http://dx.doi.org/10.1080/23270012.2016.1265906.

[43] M. Aria, C. Cuccurullo, Bibliometrix: An r-tool for comprehensive science mapping analysis, J. Inf. 11 (2017) 959–975, http://dx.doi.org/10.1016/j.joi.2017.08.007.

[44] M. Gutierrez-Bustamante, L. Espinosa-Leal, Natural language processing methods for scoring sustainability reports—a study of nordic listed companies, Sustainability 14 (9165) (2022) http://dx.doi.org/10.3390/su14159165.

[45] B. Koloski, S. Montariol, M. Purver, S. Pollak, Knowledge informed sustainability detection from short financial texts, in: Proc. 4th Workshop Fintech Nat. Lang. Process., FinNLP, 2022, pp. 228–234, http://dx.doi.org/10.18653/v1/2022.finnlp-1.31.

[46] H. Kang, J. Kim, Analyzing and visualizing text information in corporate sustainability reports using natural language processing methods, Appl. Sci. 12 (5614) (2022) http://dx.doi.org/10.3390/app12115614.

[47] J.A. Bingler, M. Kraus, M. Leippold, N. Webersinke, Cheap talk and cherry-picking: What climatebert has to say on corporate climate risk disclosures, Financ. Res. Lett. 47 (2022) 102776, http://dx.doi.org/10.1016/j.frl.2022.102776.

[48] L. Hillebrand, M. Pielka, D. Leonhard, T. Deußer, T. Dilmaghani, B. Kliem, R. Loitz, M. Morad, C. Temath, T. Bell, et al., Sustain.ai: a recommender system to analyze sustainability reports, in: Proc. Nineteenth Int. Conf. Artif. Intell. Law, 2023, pp. 412–416, http://dx.doi.org/10.1145/3594536.3595131.

[49] A. Kazakov, S. Denisova, I. Barsola, E. Kalugina, I. Molchanova, I. Egorov, A. Kosterina, E. Tereshchenko, L. Shutikhina, I. Doroshchenko, et al., Esgify: Automated classification of environmental, social, and corporate governance risks, in: Doklady Mathematics, Springer, 2023, pp. S529–S540, http://dx.doi.org/10.1134/S1064562423701673.

[50] G. Morio, S.Y. In, J. Yoon, H. Rowlands, C. Manning, Reportparse: A unified nlp tool for extracting document structure and semantics of corporate sustainability reporting, in: Proc. 33rd Int. Joint Conf. Artif. Intell. International Joint Conferences on Artificial Intelligence Organization, 2024, pp. 8749–8753, http://dx.doi.org/10.24963/ijcai.2024/1024.

[51] Y. Li, M. Rockinger, Unfolding the transitions in sustainability reporting, Sustainability 16 (809) (2024) http://dx.doi.org/10.3390/su16020809.

[52] T. Schimanski, A. Reding, N. Reding, J. Bingler, M. Kraus, M. Leippold, Bridging the gap in esg measurement: Using nlp to quantify environmental, social, and governance communication, Financ. Res. Lett. 61 (2024) 104979, http://dx.doi.org/10.1016/j.frl.2024.104979.

[53] H.A. Long, D.P. French, J.M. Brooks, Optimising the value of the critical appraisal skills programme (casp) tool for quality appraisal in qualitative evidence synthesis, Res. Methods Med. Health Sci. 1 (2020) 31–42, http://dx.doi.org/10.1177/2632084320947559.

[54] T. Nugent, N. Stelea, J.L. Leidner, Detecting environmental, social and governance (esg) topics using domain-specific language models and data augmentation, in: Flexible Query Answering Systems: 14th International Conference, FQAS 2021, Bratislava, Slovakia, September (2021) 19–24, Proceedings 14, Springer, 2021, pp. 157–169, http://dx.doi.org/10.1007/978-3-030-86967-0_12.

[55] R. Jaiswal, S. Gupta, A.K. Tiwari, Decoding mood of the twitterverse on esg investing: Opinion mining and key themes using machine learning, Manag. Res. Rev. (2024) http://dx.doi.org/10.1108/MRR-07-2023-0526.

[56] J. Fischbach, M. Adam, V. Dzhagatspanyan, D. Mendez, J. Frattini, O. Kosenkov, P. Elahidoost, Automatic esg assessment of companies by mining and evaluating media coverage data: Nlp approach and tool, in: 2023 IEEE International Conference on Big Data (BigData), IEEE, 2023, pp. 2823–2830, http://dx.doi.org/10.1109/BigData59044.2023.10386488.

[57] K. Nakagawa, S. Sashida, R. Kitajima, H. Sakai, What do good integrated reports tell us?: An empirical study of japanese companies using text-mining, in: 2020 9th International Congress on Advanced Applied Informatics, IIAI-AAI, IEEE, 2020, pp. 516–521, http://dx.doi.org/10.1109/IIAI-AAI50415.

[58] D. Wang, N. Raman, M. Sibue, Z. Ma, P. Babkin, S. Kaur, Y. Pei, A. Nourbakhsh, X. Liu, Docllm: A layout-aware generative language model for multimodal document understanding, 2023, https://arxiv.org/abs/2401.00908, arXiv:2401.00908.

[59] D. Croce, G. Castellucci, R. Basili, GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2020, pp. 2114–2119, http://dx.doi.org/10.18653/v1/2020.acl-main.191, Online, https://aclanthology.org/2020.acl-main.191.

[60] Z. Shi, F. Tonolini, N. Aletras, E. Yilmaz, G. Kazai, Y. Jiao, Rethinking semi-supervised learning with language models, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Findings of the Association for Computational Linguistics: ACL 2023, Association for Computational Linguistics, Toronto, Canada, 2023, pp. 5614–5634, http://dx.doi.org/10.18653/v1/2023.findings-acl.347, https://aclanthology.org/2023.findings-acl.347.

[61] M. Palmer, D. Gildea, N. Xue, Semantic Role Labeling, first ed., Morgan and Claypool Publishers, 2010.

[62] D. Lusseau, F. Mancini, The sustainome of global goal interactions varies by country income and is disproportionately influenced by inequalities, 2018, http://dx.doi.org/10.48550/arXiv.1804.09095, arXiv preprint arXiv:1804.09095.

[63] L. Carlson, S.J. Grove, N. Kangun, A content analysis of environmental advertising claims: A matrix method approach, J. Advert. 22 (1993) 27–39, http://dx.doi.org/10.1080/00913367.1993.10673409.

[64] Z. Chen, M. Wu, A. Chan, X. Li, Y.S. Ong, Survey on ai sustainability: Emerging trends on learning algorithms and research challenges, IEEE Comput. Intell. Mag. 18 (2023) 60–77, http://dx.doi.org/10.1109/MCI.2023.3245733.

[65] T. Vaiyapuri, S.K. Jagannathan, M.A. Ahmed, K.C. Ramya, G.P. Joshi, S. Lee, G. Lee, Sustainable artificial intelligence-based twitter sentiment analysis on covid-19 pandemic, Sustain 15 (6404) (2023) http://dx.doi.org/10.3390/su15086404.

[66] C. Gaudet-Blavignac, J.L. Raisaro, V. Touré, S. Österle, K. Crameri, C. Lovis, A national, semantic-driven, three-pillar strategy to enable health data secondary usage interoperability for research within the swiss personalized health network: Methodological study, JMIR Med. Inf. 9 (2021) e27591, http://dx.doi.org/10.2196/27591.

[67] Y. Fukuoka, T.G. Lindgren, Y.D. Mintz, J. Hooper, A. Aswani, Applying natural language processing to understand motivational profiles for maintaining physical activity after a mobile app and accelerometer-based intervention: The mped randomized controlled trial, JMIR MHealth UHealth 6 (2018) e10042, http://dx.doi.org/10.2196/10042.

[68] M.Z.F. Khairuddin, S. Sankaranarayanan, K. Hasikin, N.A. Abd Razak, R. Omar, Contextualizing injury severity from occupational accident reports using an optimized deep learning prediction model, PeerJ Comput. Sci. 10 (2024) e1985, http://dx.doi.org/10.7717/peerj-cs.1985.

[69] M.J. Usigbe, S. Asem-Hiablie, D.D. Uyeh, O. Iyiola, T. Park, R. Mallipeddi, Enhancing resilience in agricultural production systems with ai-based technologies, Env. Dev. Sustain. 1 (2023) 1–29, http://dx.doi.org/10.1007/s10668-023-03588-0.

[70] N. Smeuninx, B. De Clerck, W. Aerts, Measuring the readability of sustainability reports: A corpus-based analysis through standard formulae and nlp, Int. J. Bus. Commun. 57 (2020) 52–85, http://dx.doi.org/10.1177/2329488416675456.

[71] C.P. Almendros, L.E. Anke, R.T. Valdés, Corporate social responsibility in the ibex 35: A corpus-based study of csr reports, Procedia-Soc. Behav. Sci. 95 (2013) 612–620, http://dx.doi.org/10.1016/j.sbspro.2013.10.689.

[72] C.A. Adams, G.R. Frost, Integrating sustainability reporting into management practices, Acc. Forum 32 (2008) 288–302, http://dx.doi.org/10.1016/j.accfor.2008.05.002.

[73] S. Hamza, A. Jarboui, Esg performance and disclosure quality: Does a well-balanced board of directors matter when impression management occurs? EuroMed J. Bus. (2024) http://dx.doi.org/10.1108/EMJB-05-2023-0140.

[74] N. Székely, J. Vom Brocke, What can we learn from corporate sustainability reporting? deriving propositions for research and practice from over 9 500corporate sustainability reports published between 1999 and 2015 using topic modelling technique, PLoS One 12 (2017) e0174807, http://dx.doi.org/10.1371/journal.pone.0174807.

[75] R. Niveditha, N.K. NS, M.R. Parimi, A. Raam, S. Babu, Develop csr themes using text-mining and topic modelling techniques, in: 2020 IEEE Int. Conf. Cloud Comput. Emerg. Markets, CCEM, IEEE, 2020, pp. 67–71, http://dx.doi.org/10.1109/CCEM50674.2020.00024.

[76] D.M. Blei, Probabilistic topic models, Commun. ACM 55 (2012) 77–84, http://dx.doi.org/10.1145/2133806.2133826.

[77] Y. Nakao, A. Ishino, K. Kokubu, H. Okada, Exploring visual communication in corporate sustainability reporting: Using image recognition with deep learning, Corp. Soc. Responsib. Env. Manag. (2024) http://dx.doi.org/10.1002/csr.2735.

[78] F. Manservisi, M. Banzi, T. Tonelli, P. Veronesi, S. Ricci, D. Distante, S. Faralli, G. Bortone, Environmental complaint insights through text mining based on the driver, pressure, state, impact, and response (dpsir) framework: Evidence from an italian environmental agency, Reg. Sustain. 4 (2023) 261–281, http://dx.doi.org/10.1016/j.regsus.2023.08.002.

[79] S.J. Rivera, B.S. Minsker, D.B. Work, D. Roth, A text mining framework for advancing sustainability indicators, Env. Model. Softw. 62 (2014) 128–138, http://dx.doi.org/10.1016/j.envsoft.2014.08.016.

[80] M. Geissdoerfer, P. Savaget, N.M.P. Bocken, E.J. Hultink, The circular economy–a new sustainability paradigm? J. Clean. Prod. 143 (2017) 757–768, http://dx.doi.org/10.1016/j.jclepro.2016.12.048.

[81] P. Velte, Automated text analyses of sustainability & integrated reporting. a literature review of empirical-quantitative research, J. Glob. Responsib. 14 (2023) 530–566, http://dx.doi.org/10.1108/JGR-09-2022-0090.

[82] M. Castelo Branco, The new wave of csr-related mainstream accounting research: On the performative effects of literature reviews, Soc. Env. Accountab. J. 44 (2024) 37–62, http://dx.doi.org/10.1080/0969160X.2024.2319882.

[83] C. Cunha, L. Duncan, Integrated reporting and group decision-making for sustainable development: A bibliometric review, Eur. J. Sustain. Dev. 13 (2024) http://dx.doi.org/10.14207/ejsd.2024.v13n1p389, 389–389.