Information Fusion 115 (2025) 102726

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Explainable natural language processing for corporate sustainability analysis

Keane Ong [a,b], Rui Mao [c], Ranjan Satapathy [d], Ricardo Shirota Filho [d], Erik Cambria [c], Johan Sulaeman [b,e,f], Gianmarco Mengaldo [a,b,e,g,*]

[a] College of Design and Engineering, National University of Singapore, 9 Engineering Drive 1, 117575, Singapore
[b] Asian Institute of Digital Finance, National University of Singapore, Innovation 4.0, 3 Research Link, #04-03, 117602, Singapore
[c] College of Computing and Data Science, Nanyang Technological University, 50 Nanyang Ave, 639798, Singapore
[d] Institute of High Performance Computing, Agency for Science, Technology and Research, Fusionopolis Way, #16-16 Connexis, 138632, Singapore
[e] Sustainable and Green Finance Institute, National University of Singapore, Innovation 4.0, 3 Research Link, #02-02, 117602, Singapore
[f] NUS Business School, National University of Singapore, Mochtar Riady Building, 15 Kent Ridge Dr, Mocthar Riady Building, 119245, Singapore
[g] Honorary Research Fellow, Imperial College London, United Kingdom

ARTICLE INFO

Keywords:
Sustainability analysis
Corporate sustainability
Sustainability disclosure
Explainable artificial intelligence
Explainable natural language processing

ABSTRACT

Sustainability commonly refers to entities, such as individuals, companies, and institutions, having a non-detrimental (or even positive) impact on the environment, society, and the economy. With sustainability becoming a synonym of acceptable and legitimate behaviour, it is being increasingly demanded and regulated. Several frameworks and standards have been proposed to measure the sustainability impact of corporations, including United Nations' sustainable development goals and the recently introduced global sustainability reporting framework, amongst others. However, the concept of corporate sustainability is complex due to the diverse and intricate nature of firm operations (*i.e.* geography, size, business activities, interlinks with other stakeholders). As a result, corporate sustainability assessments are plagued by subjectivity both within data that reflect corporate sustainability efforts (*i.e.* corporate sustainability disclosures) and the analysts evaluating them. This subjectivity can be distilled into distinct challenges, such as incompleteness, ambiguity, unreliability and sophistication on the data dimension, as well as limited resources and potential bias on the analyst dimension. Put together, subjectivity hinders effective cost attribution to entities non-compliant with prevailing sustainability expectations, potentially rendering sustainability efforts and its associated regulations futile. To this end, we argue that Explainable Natural Language Processing (XNLP) can significantly enhance corporate sustainability analysis. Specifically, linguistic understanding algorithms (lexical, semantic, syntactic), integrated with XAI capabilities (interpretability, explainability, faithfulness), can bridge gaps in analyst resources and mitigate subjectivity problems within data.

1. Introduction

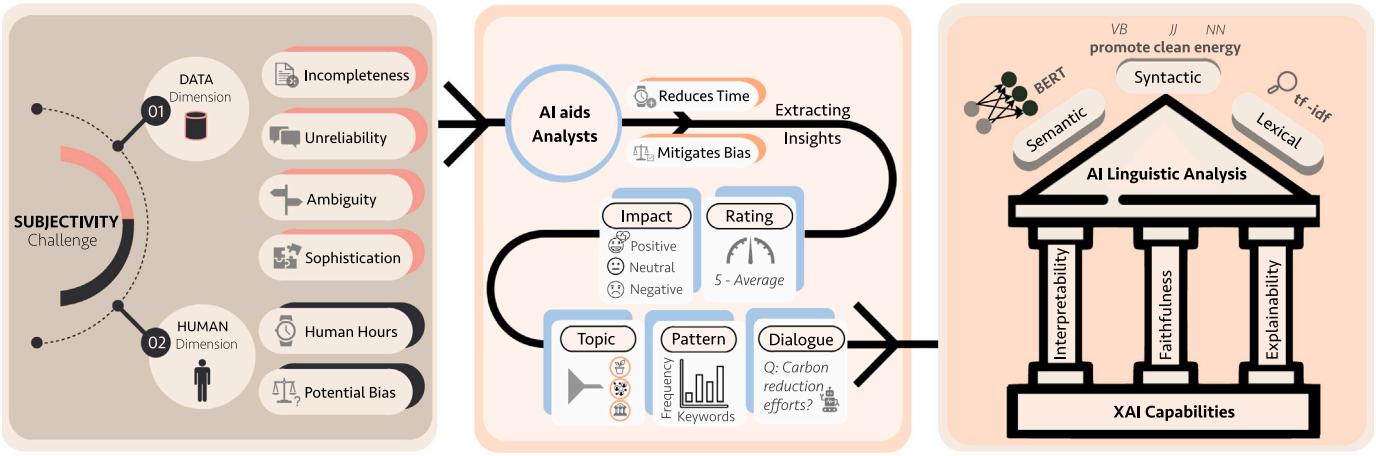
Sustainability, intended as having a non-detrimental impact on the environment, society, and the economy, is becoming increasingly essential for humanity's future. Various efforts are being undertaken at different levels of the societal and economic hierarchy, from country-level to institutional- and business-level entities, and in some cases, down to individuals. These efforts are focused on making these entities *sustainable*, amid reputational risks [1], and pressing challenges including global climate change, widespread inequity, governance malpractices, and geopolitical instability [2]. In this work, we focus on sustainability analyses of entities that provide sustainability reports (also referred to as sustainability disclosures); in particular public companies and institutions.

These entities are crucial components in the sustainable development [3], and we refer to their sustainability as *corporate sustainability*. Given that these entities have a substantial stake in global sustainability, the analysis of their sustainability, or corporate sustainability analysis, is critical [4]. Yet, this is extremely *challenging* due to the inherent complexity of corporate sustainability as a concept.

To elaborate, the complexities within firms (*e.g.*, geography, size, business activities) and outside the firms (*e.g.*, their global supply chains), as well as the firms' relationships with non-business stakeholders, lead to the evolving frameworks and guidelines put forward to address corporate sustainability issues [5,6]. Further complicating the matter is the lack of globally mandated standards for sustainability reporting [7].

* Corresponding author.
E-mail address: mpegim@nus.edu.sg (G. Mengaldo).

https://doi.org/10.1016/j.inffus.2024.102726
Received 31 July 2024; Received in revised form 26 September 2024; Accepted 2 October 2024
Available online 11 October 2024



A. Subjectivity leads to data and human challenges B. AI addresses subjectivity by extracting insights C. Strategies for effective XNLP deployment

Fig. 1. Proposed framework for XNLP enhanced sustainability analysis.

For example, a firm may use a convenient corporate sustainability reporting framework that magnifies only certain aspects of a firm's practices [8], while omitting key sustainability dimensions [9]. Or, within a given regulatory disclosure framework that a firm shall abide to, it may provide sustainability disclosures that are difficult for external parties to digest, and whose integrity and transparency may be questionable [10]. This first aspect, that we label the *data dimension*, is inherently *subjective*, as it depends on what the firm chooses to disclose and how. However, this data dimension is only one side of the challenge.

Indeed, today's corporate sustainability analysis is carried out by human analysts who read sustainability disclosures provided by companies (*i.e.*, the data dimension), and provide corporate sustainability evaluations based on a combination of sustainability frameworks (*e.g.*, Environmental, Social, Governance (ESG) [11], Global Reporting Initiative (GRI) [12], Sustainability Accounting Standards Board (SASB) [13], Greenhouse Gas Protocol and Carbon Disclosure Project (CDP) [14], United Nations' Sustainable Development Goals (SDG) [15], United Nation Global Compact (UNGC) [16]). These evaluations require an extensive amount of human hours [17], and yet are inevitably influenced by the inherent biases in the frameworks adopted as well as the analysts' own subjectivity [18,19].

This second aspect, that we label the *analyst dimension*, is also inherently *subjective*, similar to the *data dimension*. Amid the complexities of corporate sustainability, analysts face the challenge of dealing with *subjective* (and likely flawed) data, and they need to account for their own *subjective* perspective. The interplay between the *data* and *analyst* dimensions is critical, as it results in a highly subjective evaluation of corporate sustainability that would hamper efforts to achieve corporate sustainability goals.

We argue that, to address the subjectivity inherent to data and analysts, it is crucial to adopt artificial intelligence (AI), specifically through linguistic understanding methods enhanced by explainable AI (XAI) capabilities, or what is termed as explainable natural language processing (XNLP) [20]. To qualify, adopting XNLP entails *complementing* human analysts instead of replacing them. Within corporate sustainability analysis, analysts should remain 'in the loop' — while XNLP can expeditiously process numerous sustainability disclosures to unveil consistent and meaningful insights, analysts can further synergise these insights to inform their corporate sustainability assessments.

In this paper, we follow the three foundational building blocks presented in Fig. 1, focusing on how XNLP can pave the way for effective corporate sustainability analysis by addressing the *subjectivity* issues that plague the field. We first break down the subjectivity challenge into its *data* and *analysts* dimensions (Fig. 1(A)), characterising the problems for XNLP to address. Next, we consolidate NLP's distinct advantages for solving these challenges by delineating its specialised applications (Fig. 1(B)). Finally, we lay the groundwork for effective XNLP deployment within this domain (Fig. 1(C)). We realise this by surveying NLP methods to match specific applications based on their different levels of linguistic understanding, and proposing strategies to integrate XAI capabilities. By offering insights that blend technical depth with broader viewpoints of sustainability, we pave the way for XNLP's adoption within corporate sustainability analysis.

2. The subjectivity challenge

As a result of the *complexity* inherent to corporate sustainability, *subjectivity* commonly underpins corporate sustainability analysis. Specific challenges arising from subjectivity can be distilled into the *data* and *analyst* dimensions. We expound on these dimensions to frame (X)NLP's suitability for corporate sustainability analysis.

2.1. The data dimension

Data is the first building block, or entry point, of corporate sustainability analysis, as depicted in (Fig. 1(A)). While the data required for corporate sustainability assessments can come from various sources, *corporate sustainability disclosures* are a predominant source [21]. These disclosures consist of reports released by businesses detailing their own sustainability efforts [21]. For example, *general sustainability reports* typically follow established sustainability frameworks to describe a company's sustainability efforts and achievements [22], and are typically accompanied by *integrated reports* that aggregate the general sustainability information with financial implications [23].

Additional data may come from *government-mandated sustainability disclosures*, where companies are required to release specific aspects of their sustainability performance [24], such as their environmental impact [25], or ESG (Environmental, Social, and Governance) practices [26]. Data originating from NGO sources and media coverage may also be relevant to assess a company's sustainability efforts [27,28]. Yet these data are influenced by sustainability disclosures [29,30], underlining the importance of focusing our discussion on the sustainability disclosures themselves. Due to the intricacies and multi-faceted dimensions of corporate sustainability, these data sources inherently contain *subjectivity*. They involve competing notions of corporate sustainability, expressed through a vast and diverse array of sustainability frameworks [31,32].

Moreover, even within a specific framework, deciding what information to convey necessitates a value judgement (*e.g.*, financial vs. environmental materiality), further exacerbating the partiality involved [33]. We argue that this *subjectivity* at the data level manifests itself in four data grand challenges that analysts need to face, namely *incompleteness*, *unreliability*, *ambiguity*, and *sophistication*. These 4 data challenges are reported on the first block of (Fig. 1(A)), referred to as *data dimension*. We detail them in the following.

Incompleteness. It refers to how data may not provide a comprehensive and complete view of a firm's sustainability efforts. As certain sustainability disclosures are voluntary, companies can decide on what is material enough to release as sustainability information, thereby raising concerns about openness and the omission of important data [33].

Unreliability. It refers to the trustworthiness and accuracy of sustainability data. Indeed, corporate and government-linked sustainability disclosures may involve greenwashing, where organisations misleadingly claim and exaggerate that their activities or products are more environmentally friendly than they really are [34]. Or they use non-transparent carbon offset products, that often do not deliver the offsets promised [35]. In fact, such developments are particularly frequent for firms that are of larger sizes, to contend with increased dealing with stakeholders [36].

Ambiguity. It refers to data being vague and unclear, making negative content less conspicuous [37]. One example is sharing information without accompanying it with the appropriate context [38]. This may lead analysts to misunderstand the sustainability efforts of companies. Ambiguity is related, to some extent, to greenwashing [39], although it does not explicitly refer to inaccurate or misleading data as unreliability does.

Sophistication. It refers to sustainability disclosures being tediously lengthy, text-heavy, and requiring specialised knowledge to comprehend, making them an extremely sophisticated writing category [40]. As a result, firm sustainability reports can require many hours of human effort to understand, let alone derive useful insights [41].

2.2. The analyst dimension

Human analysts must synthesise and interpret sustainability data to produce accurate corporate sustainability assessments. To describe this process, relevant data is collected according to the objectives and scope of the corporate sustainability analysis (*i.e.* evaluating all operations versus one area). Thereafter, the analysis entails leveraging sustainability frameworks or metrics (*i.e.*, carbon footprint) to measure different impact dimensions (environmental, social, economic etc.) [42]. The costs and inaccuracies of this process are exacerbated by human limitations. Specifically, human analysts cannot deal well with the challenges of *subjective* data, and are not free from *subjective* interpretations. These human-centric issues are encapsulated by *limited human-hours* and *potential bias*.

Limited human-hours. Analysts have a limited amount of time to produce corporate sustainability assessments. However, the process entails extracting meaningful information from corporate sustainability disclosures [41], which requires a significant amount of time *i.e.*, human-hours [17]. In particular, the disclosures are extremely time consuming to read and understand due to the *sophistication* and *ambiguity* of the data within them. Moreover, given the potential *incompleteness* of disclosure information, analysts may analyse additional sources to ascertain a firm's sustainability, further increasing the human-hours required [43].

Potential Bias. Analysts' interpretations of sustainability data are subject to bias. This is not least due to the complex and wide ranging factors to be considered, such as firm characteristics, evolving sustainability guidelines, and economic implications, amongst others, often leading to varying interpretations amongst analysts [18,19]. To further compound the potential for bias, data *ambiguity* and *unreliability* hinders clear interpretation of sustainability information by obfuscating negative content [37].

2.3. NLP to the rescue

Subjectivity pervades through the data and analyst dimensions, making corporate sustainability analyses an onerous task. To alleviate this burden, we argue that AI technologies, through linguistic understanding enabled by natural language processing (NLP), can significantly aid analysts [44,45]. NLP can automatically process hundreds of sustainability disclosures [46], summarising key details and extracting insights for analysts. This reduces the *human hours* required to analyse disclosures and partially addresses *potential bias* by providing more consistent and replicable insights [47,48]. We concretise NLP's usefulness in this domain by detailing five NLP tasks for corporate sustainability analysis **(T1–T5)** that we deem critical, and that have been partially explored in the literature.

(T1) Topic extraction. Topic extraction allows an analyst to obtain, from sustainability disclosures, content most pertinent to corporate sustainability. Consequently, analysts do not need to sieve through all information, and can focus their analysis on material information [49]. *Thematic* extraction involves classifying sentences within sustainability frameworks. This includes sorting text into GRI topics [50], ESG pillars [51], or ESG-related concepts [52]. *Universal* extraction filters textual data into generic categories that are separate from sustainability frameworks. For instance, labelling environmental claims [46], emissions targets [53], or climate relevant information [49]. On the other hand, *topic discovery* determines topics from sustainability disclosures a posteriori. For example, sustainability topics were uncovered from the sustainability disclosures of shipping companies [54].

(T2) Impact Classification. Impact insights can also be derived from sustainability disclosures, providing analysts a valuable perspective on corporate sustainability. *Polarity* classification determines the impact polarity of text from sustainability disclosures [55]. For instance, within SEC 10k reports, sustainability related sentences can be labelled as positive, negative or neutral in relation to their impact toward sustainability [56]. Alternatively, *strategic* classification uncovers a text's prospective impact. Examples include the risk and opportunity impact analysis of ESG-related texts [57].

(T3) Rating. A firm's sustainability performance can be automatically scored to expedite an analyst's evaluation process. *Aggregated* rating methods include computing the frequency of sentiment labels for ESG-related headlines [58], and averaging the E, S, and G topic classification probabilities of documents [59]. *Direct* methods explicitly score companies without amalgamating already labelled data. For instance, the prediction of ESG risk ratings from sustainability topics [60], as well as regressing lexical features to derive ESG risk ratings [61]. Outside of holistic sustainability frameworks (*i.e.* ESG), *direct* methods also include estimating carbon emissions from bank transactions [62], and the construction of sustainability indices from the lexical features of disclosures [63].

Table 1
Classification of NLP papers for corporate sustainability analysis, detailing the algorithm category and tasks covered.

Paper	Lexical	Semantic	Syntactic	Topic Extraction			Impact Classification		Rating		Dialogue	Linguistic Patterns
				Thematic	Universal	Topic Discovery	Polarity	Strategic	Aggregated	Direct		
[50]	✓	✓		✓			✓					
[51]		✓		✓								
[46]	✓	✓			✓							
[49,53]		✓			✓							
[54]	✓					✓						
[52]	✓	✓	✓	✓	✓							
[56]	✓	✓					✓					
[57]		✓		✓				✓				
[58]	✓	✓		✓	✓		✓		✓			
[59]		✓		✓				✓	✓			
[60]	✓	✓				✓				✓		
[62]		✓	✓							✓		
[68]		✓								✓		
[41]	✓	✓		✓								
[66]	✓											✓
[67]	✓		✓									✓
[17,64]		✓									✓	
[61,63]	✓									✓		

(T4) Dialogue. Processing sustainability disclosures into interactive dialogue format allows an analyst a user-friendly way for querying information. For instance, a chatbot integrated with Retrieval-augmented Generation (RAG) can answer analysts' questions about disclosures, allowing them to extract information without reading the copious amounts of text typically contained within them [17,64]. By emphasising critical information, dialogue systems mitigate the obfuscation of key information within disclosures, potentially enhancing the accuracy of sustainability analysis [65].

(T5) Linguistic Patterns. Linguistic analysis deepens an analyst's understanding of how a disclosure's style and structure relates with corporate sustainability performance. For example, keyword frequency and word relationships within oil and gas compliance reports can uncover environmental violation patterns [66], and the syntactic complexity of CSR reports for good CSR performers can also be distinguished [67].

3. NLP methods for linguistic understanding

Tasks T1–T5 help human analysts decode sustainability disclosures, addressing the domain's *subjectivity* challenge. To successfully accomplish T1–T5, it is useful to comprehend NLP methods according to their different levels of *linguistic understanding* (lexical, semantic, and syntactic). These three categories unpack the complexity of NLP algorithms, from simple lexical evaluation to complex semantic and syntactic processing, providing insight into the suitability of certain methods for specific tasks. To this end, Table 1 presents a structured overview, detailing which tasks, T1–T5, have been addressed in the literature by the three categories of NLP methods introduced. We delve more into these three different approaches, providing a concise review of what has been already done in the corporate sustainability analysis space.

3.1. Lexical methods

Lexical-based methods focus on the statistical occurrence of keywords and terms within sustainability data, thereby being relatively simple and easy to interpret.

More specifically, lexical methods identify lexical terms appearing in text through string matching, and compute their associated frequency (or rate) of occurrence. A formalisation of this framework is given by the term frequency–inverse document frequency (TF–IDF) equations

$$\text{TF–IDF}(t, d, D) = \text{TF}(t, d) \cdot \text{IDF}(t, D), \tag{1a}$$

$$\text{TF}(t, d) = \frac{f_{t,d}}{\sum_{l' \in d} f_{l',d}}, \tag{1b}$$

$$\text{IDF}(t, D) = \log \frac{N}{\{d \in D : t \in d\}}, \tag{1c}$$

where TF is the term frequency that quantifies the rate of occurrence of a lexical term t in a document d (*e.g.*, sustainability disclosures), computed as a proportion of all terms in the document — *i.e.*, Eq. 1b, while IDF is the inverse document frequency that measures the importance of a term across multiple documents Eq. 1c. A lexical term is less important if it appears in more documents, and vice-versa. This prevents words which are commonly used, but have relatively insignificant meaning such as 'he, she, they', to be flagged as significant. The quantity TF–IDF in Eq. 1a provides a measure of the overall statistical significance of a keyword, by weighting Eq. 1b & 1c. The three quantities, TF, IDF, and TF–IDF may be utilised independently from one another, subject to the use case.

Given their simplicity, lexical methods are appropriate for relatively uncomplicated tasks that do not require significant linguistic understanding. For example, lexical methods can be deployed for task (T5) *linguistic patterns*, to derive keyword (term) frequency patterns in sustainability disclosures [66]. On the other hand, for more complex

tasks such as T1–T4, lexical analysis is utilised as a feature engineering tool rather than a standalone method. For example, TF–IDF vectorisation has been employed to identify companies of interest for ESG classification from news headlines [58], while TF–IDF-extracted features have been used to detect environmental claims [46].

3.2. Semantic methods

Semantic methods interpret textual content to understand its meaning [69]. This can extend beyond understanding the presence of individual words, to involve contextual understanding and implicit connotations. Unlike lexical methods that simply identify word occurrences, semantic approaches often leverage word embeddings, which are vector representations of words within a multi-dimensional space. We further distinguish these two approaches by analysing the sentence "*Renewable energy is sustainable, as it does not utilise finite resources*". A lexical approach registers the occurrences of individual words (*i.e.* 'renewable', 'sustainability', 'finite'), without grasping their deeper and interconnected meanings. On the other hand, semantic methods, by determining the proximity of their respective word vectors, interpret 'renewable' as synonymous with 'sustainability' but antonymous with 'finite'. This helps semantic algorithms comprehend that 'renewable' is conceptually linked to 'sustainability', unlike 'finite' which implies that resources are exhaustible.

Traditional methods such as GloVe derive text embeddings from word co-occurrences, and can be leveraged to capture semantic relationships [70]. For instance, similarity methods can be employed on word embeddings to classify sentences according to their relevance to each ESG indicator [41]. However, while useful, these traditional methods may not capture linguistic subtleties comprehensively, paving the way for more advanced techniques.

As a significant leap in semantic understanding, deep learning models such as BERT [71] powerfully capture the intricacies of text. BERT employs bidirectional attention mechanisms to compute the two-way influence between words. As a result, important contextual and semantic information of a sentence can be captured. BERT develops general language understanding through pre-training on a wide variety of text corpora, and subsequently is fine tuned for specific tasks.

As they attain a high level of semantic understanding, BERT models are suitable for tasks within the corporate sustainability analysis field that require complex language understanding, without being specifically structured for generative tasks. Therefore, BERT is appropriate for complex language understanding tasks like (T1) *topic extraction*, (T2) *impact classification* and (T3) *rating*, but not generative tasks such as (T4) *dialogue*. In line with this, BERT-like models have been used for ESG label classification [51], producing company environmental scores [68], and to derive topics and their associated keywords for ESG-related risk factors [60] using BERT-generated text embeddings.

Similar to BERT, generative large language models (Gen-LLMs) such as GPT-4 also involve the comprehension of text by exploiting deep learning based attention architectures [72], and pre-training on large text corpora. However, by learning to predict the next word given the preceding text [73], their focus extends beyond understanding language, to the generation of coherent text, as well as inference and reasoning capabilities [74]. This makes them well-suited for generative answering and summarisation tasks found within (T4) *dialogue*, and advantageous for reasoning tasks like (T3) *rating* which involves complex evaluation [75]. As such, generative LLMs have been leveraged for transforming TCFD disclosures into dialogue format, as well as evaluating sustainability disclosures for conformity to TCFD reporting guidelines [17]. The latter demonstrates the reasoning abilities of generative LLMs, showing how they can be leveraged to evaluate corporate sustainability performance from disclosures.

3.3. Syntactic methods

Syntactic approaches involve analysing the structure of sentences in terms of grammatical and language rules [76]. This contrasts with other approaches like analysing the occurrence of lexicons and interpreting textual meaning. To elaborate on the latter, while semantic methods can involve a higher level of abstraction (*i.e.*, comprehending the hidden or implied connotations of words), syntactic measures focus on the organisation of words and phrases within a sentence without interpreting them more broadly.

Syntactic methods include part-of-speech (POS) tagging [77], dependency parsing [78], and constituency parsing [79], amongst others. These methods can augment efforts to accomplish several complex NLP tasks for corporate sustainability analysis, although they have only seen limited application so far. For instance, when combined with other methods like semantic analysis, text can be analysed more granularly. In other domains such as finance and economics, aspect-based sentiment analysis (ABSA) has already been achieved through POS tagging and semantic rules [80,81]. Such integrated approaches can more intricately tackle task (T2) *impact classification*, by revealing impact insights toward aspects. To illustrate with an example, within the sentence *"Company X has good emissions performance despite having poor employee wellness"*, ABSA allows for the aspect 'emissions performance' to be labelled as positive and the aspect 'employee wellness' to be labelled as negative. Distinguishing between the impact on the two aspects enables a finer analysis of sustainability text. Additionally, syntactic analysis can also be used for identifying companies and their relationships with respect to sustainability activities [82]. Such methodologies can be enabled by explicit syntactic attributes (*i.e.*, POS, dependency parsing) in addition to semantic patterns. Topics can be derived from the extracted links between companies and their sustainability practices to solve task (T1) *topic extraction*. Additionally, these links can be further processed to evaluate corporate sustainability performance, in the spirit of task (T3) *rating*. Other applications of syntactic analysis include evaluating the grammatical clauses of disclosure sentences [67], in line with (T5) *linguistic patterns*, or the parsing of sustainability-related concepts, in a similar vein to the extraction of financial concepts [83].

4. XNLP enhances subjectivity mitigation

The exploration of NLP methods from the *lexical, semantic, syntactic* perspectives provides a pathway for achieving tasks T1–T5, mitigating the *subjectivity* challenges within sustainability disclosures. As we move forward, our focus shifts toward enhancing these efforts by scaffolding the NLP methods with XAI capabilities (*interpretability, explainability and faithfulness*), forming the core components of explainable natural language processing (XNLP) as defined earlier. Despite their potential, these XAI features are still uncommon within NLP for corporate sustainability analysis, representing a promising research area that has yet to gain traction. We describe these capabilities below.

Interpretability. An AI model's capacity to provide an understanding of its mechanism [84]. For example, scoring which keywords are most salient for an algorithm's classification [85].

Explainability. An AI model's capacity to explain why it produces an output [84]. For example, explaining the reasoning steps for model decisions, as enabled by large language models [86].

Faithfulness. The accuracy of an AI model's *interpretability* and *explainability* with respect to its true workings [87–89]. For example, the extent to which a model's feature salience scores and provided explanations are representative of its actual workings. Analysing *faithfulness* can be challenging given the lack of ground truth for a model's explainability or interpretability, with several approaches involving dataset modification to verify the feature importance estimated by interpretability methods [90].

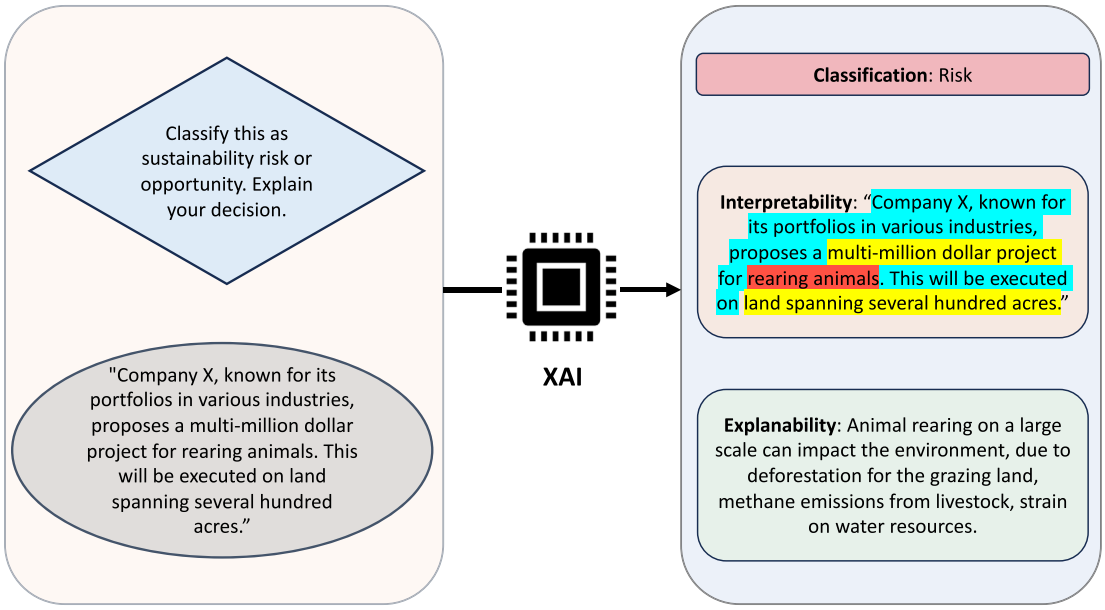


Fig. 2. Alongside outputting a decision for sustainability impact classification, XAI can provide an interpretation and explanation.

Integrating XAI capabilities extends the effectiveness of NLP methods, ensuring they further address the *subjectivity* challenges of the data and analyst dimensions. This ultimately enhances NLP-driven corporate sustainability analysis. We develop this idea by elaborating on the value of these XAI capabilities for NLP-driven corporate sustainability analysis.

4.1. Achieving trustworthy insights

XAI capabilities allow NLP models to extend beyond classifications and predictions, providing analysts useful insights through faithful explanations and interpretations [91,92]. These insights enhance NLP trustworthiness within corporate sustainability analysis, by clarifying model decisions along with the *ambiguity* of sustainability disclosures.

For example, Fig. 2 highlights how *interpretability, explainability* and *faithfulness* enhances trust in an NLP model's classification for sustainability text. *Interpretability* allows an analyst to verify that appropriate and relevant text features (*i.e.* 'rearing animals' instead of 'portfolios'), are exploited for classifying the sentence's risk impact (similar to the concepts used in self-interpretable image classification — *e.g.*, [93]). *Explainability* explicates the relevance of 'rearing animals' to sustainability risk, further justifying the model's classification. It explains the link between 'rearing animals', deforestation, methane emissions, and strained water resources, resolving the ambiguity surrounding the term 'rearing animals'. *Faithfulness* ensures that the insights from *interpretability* and *explainability* are meaningful in that they accurately reflect the model's mechanism. Put together, *interpretability, explainability* and *faithfulness* prove the robustness of the model's decision-making process, increasing an analyst's trust in its classification output.

Integrating these capabilities within NLP has already proven to be of significant for trust-dependent sectors like healthcare [94]. This reinforces their broader applicability to corporate sustainability analysis, where trustworthy and credible NLP insights are also critical [95]. Such integration will allow NLP models to produce more actionable insights for analysts.

4.2. Inferring patterns for greenwashing research

Interpretable NLP models can afford cues on the interdependence between different features within sustainability disclosures. This can inspire researchers to theorise and subsequently validate the causal and correlation relationships between attributes relevant to corporate sustainability analysis [91,92]. By doing so, our understanding of under-explored issues can be augmented, mitigating longstanding issues within the domain.

To elaborate, white-box models or intrinsically *interpretable* models (Naive Bayes Classifiers, Generalised Additive Models, Decision Trees etc.), allow us to grasp the decision-making process from feature inputs to classification, implying the significance of specific features and their relationships with predictions. For instance, text classification models designed with *interpretability* techniques can elucidate the influence of keywords for classification decisions [96]. Developing a similar white-box model for classifying greenwashing texts allows us to study how specific text features result in a positive classification. This can offer hints on the semantic and syntactic features that characterise greenwashed texts. From a 'machine' based perspective, it can powerfully reveal latent data patterns easy for an analyst to miss. This fresh angle would strengthen existing greenwashing research that predominantly involves human analysis of sustainability texts [34], making greenwashing more understandable and detectable. In this fashion, *interpretability* can improve the *unreliability* and *ambiguity* issues in the field arising from the greenwashing phenomenon.

4.3. Generalisability across frameworks and sources

Sustainability disclosures stem from different sources and are presented through diverse frameworks. As such, they can have differing linguistic and textual features, posing a challenge for NLP models to generalise. To elaborate, while firms can focus corporate sustainability disclosures on positive efforts [97], government-linked sustainability disclosures may mandate firms to disclose sustainability related risks (such as U.S. Securities and Exchange Commission 10k reports) [98].

An *impact classification* model trained on corporate sustainability disclosures may not adapt well to government-linked sustainability

disclosures, as the latter more frequently carries negative impacts. On top of heterogeneous sustainability data, the *generalisability* problem is further compounded by the lack of available datasets within the field [51]. This reduces the diversity of data for training robust models. In light of limited data availability, *interpretability* methods provides an alternative means for mitigating generalisability issues. Specifically, *interpretability* allows us to understand a model's sensitivity to training data, providing insight into portions of data that cause overfitting. Recent works such as [99] detail a method to do so through a memory perturbation equation. The paper faithfully derives 'shirt, pullover' as classes the model is most sensitive to while training on the FMNIST dataset, and demonstrates how removing these classes can improve model generalisability performance. Moreover, it also highlights how specific samples and training epochs impact model sensitivity. While this method has been deployed for computer vision, the same principles can be adapted to NLP for sustainability analysis. Leveraging such a tool, NLP practitioners can consider removing portions of data most prone to model overfitting. This optimises training for general performance across the different sustainability disclosures. By enhancing generalisability, NLP models can effectively analyse diverse types of information associated with a firm's sustainability efforts [51]. This potentially compensates for *incompleteness* and *unreliability* data problems, by reducing over-reliance on specific sources that omit important details or incredible information.

4.4. Mitigating bias

By producing consistent and reproducible insights, NLP can partially address the *potential bias* of analysts that conduct corporate sustainability analysis. However, NLP also risks introducing its own biases if not designed transparently. To elaborate, NLP constructed with XAI capabilities can elucidate its decision-making process, allowing analysts to verify its learned features [100]. While this may still incur the partiality of analyst judgement, an XAI enhanced model can conversely highlight potentially overlooked features to analysts, guiding them to produce more balanced corporate sustainability assessments. Consequently, XAI capabilities can enable a synergistic interaction between NLP models and analysts for reducing potential bias.

5. Conclusion

XNLP algorithms are well equipped for sustainability analysis given that they encompass both linguistic understanding and XAI capabilities. To elaborate on this, while lexical, semantic and syntactic analysis of text automates and enhances sustainability analysis, XAI further mitigates the subjectivity challenge of sustainability assessments. However, NLP for sustainability analysis has not yet embraced these directions, leaving its vast potential untapped. This underscores an opportunity for researchers and developers to explore the strategies highlighted — achieving trustworthy insights, inferring patterns for greenwashing research, generalisability across frameworks and sources, mitigating bias — to enhance NLP for tasks (T1–T5). These efforts can not only improve sustainability analysis but potentially reshape how businesses, investors, and regulators address corporate sustainability reporting and attribution.

Limitations

Given that the development of NLP methods for sustainability analysis is still nascent, several papers that describe NLP methods for sustainability analysis entail emerging research that have not yet undergone extensive peer-review. Despite our attempts to filter out papers with evident shortcomings, it is important to acknowledge that some of the discussed papers may contain non-authoritative viewpoints.

CRediT authorship contribution statement

Keane Ong: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualisation. **Rui Mao:** Writing – review & editing, Methodology, Investigation, Conceptualisation. **Ranjan Satapathy:** Investigation. **Ricardo Shirota Filho:** Writing – review & editing. **Erik Cambria:** Writing – review & editing, Conceptualisation. **Johan Sulaeman:** Writing – review & editing, Methodology, Investigation. **Gianmarco Mengaldo:** Writing – review & editing, Supervision, Methodology, Conceptualisation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research/project is supported by the NUS Sustainable and Green Finance Institute (SGFIN), NUS Asian Institute of Digital Finance (AIDF), Ministry of Education, Singapore under its MOE Academic Research Fund Tier 2 (STEM RIE2025 Award MOE-T2EP20123-0005), MOE Tier 2 Award (MOE-T2EP50221-0006: "Prediction-to-Mitigation with Digital Twins of the Earth's Weather"), MOE Tier 1 Award (MOE-T2EP50221-0028: "Discipline-Informed Neural Networks for Interpretable Time-Series Discovery"), and by the RIE2025 Industry Alignment Fund – Industry Collaboration Projects (IAF-ICP) (Award I2301E0026), administered by A*STAR, as well as supported by Alibaba Group and NTU Singapore. The authors would also like to credit Bayan Abusalameh and Jiawen Wei for designing the figures.

Data availability

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

References

[1] I. Nikolaou, K. Evangelinos, W. Leal Filho, A system dynamic approach for exploring the effects of climate change risks on firms' economic performance, J. Clean. Prod. 103 (2015) 499–506, Carbon Emissions Reduction: Policies, Technologies, Monitoring, Assessment and Modeling.
[2] Maria Folqué, Elena Escrig-Olmedo, Teresa Corzo Santamaria, Sustainable development and financial system: Integrating esg risks through sustainable investment strategies in a climate change context, Sustain. Dev. 29 (5) (2021) 876–890.
[3] Simone Cenci, Matteo Burato, Marek Rei, Maurizio Zollo, The alignment of companies' sustainability behavior and emissions with global climate targets, Nature Commun. 14 (1) (2023) 7831, 2023.
[4] R. Rajesh, Exploring the sustainability performances of firms using environmental, social, and governance scores, J. Clean. Prod. 247 (2020) 119600.
[5] Amina Buallay, Sustainability reporting and firm's performance: Comparative study between manufacturing and banking sectors, Int. J. Prod. Perform. Manag. 69 (3) (2020) 431–445.
[6] Belén Derqui, Towards sustainable development: Evolution of corporate sustainability in multinational firms, Corp. Soc. Responsib. Environ. Manag. 27 (6) (2020) 2712–2723.
[7] Florian Berg, Julian F. Kölbel, Roberto Rigobon, Aggregate confusion: The divergence of ESG ratings*, Rev. Finance 26 (6) (2022) 1315–1344.
[8] Nicole Darnall, Hyunjung Ji, Kazuyuki Iwata, Toshi H. Arimura, Do esg reporting guidelines and verifications enhance firms' information disclosure? Corp. Soc. Responsib. Environ. Manag. 29 (2022) 1214–1230.
[9] James Demastus, Nancy E. Landrum, Organizational sustainability schemes align with weak sustainability, Bus. Strategy Environ. 33 (2) (2024) 707–725.
[10] Olivier Boiral, Jean-François Henri, Is sustainability performance comparable? A study of gri reports of mining organizations, Bus. Soc. 56 (2) (2017) 283–317.
[11] Piotr Dmuchowski, Wojciech Dmuchowski, Aneta H. Baczewska-Dabrowska, Barbara Gworek, Environmental, social, and governance (esg) model; impacts and sustainable investment – global trends and poland's perspective, J. Environ. Manag. 329 (2023) 117023.

[12] Bianca Alves Almeida Machado, Lívia Cristina Pinto Dias, Alberto Fonseca, Transparency of materiality analysis in gri-based sustainability reports, Corp. Soc. Responsib. Environ. Manag. 28 (2021) 570–580.

[13] Li Li Eng, Mahelet Fikru, Thanyaluk Vichitsarawong, Comparing the informativeness of sustainability disclosures versus esg disclosure ratings, Sustain. Account. Manag. Policy J. 13 (2022) 494–518.

[14] Panayis Pitrakkos, Warren Maroun, Evaluating the quality of carbon disclosures, Sustain. Account. Manag. Policy J. 11 (2020) 553–589.

[15] Prajal Pradhan, Luís Costa, Diego Rybski, Wolfgang Lucht, Jürgen P. Kropp, A systematic study of sustainable development goal (sdg) interactions, Earth's Future 5 (11) (2017) 1169–1179.

[16] Guido Orzes, Antonella.Maria Moretto, Maling Ebrahimpour, Marco Sartor, Mattia Moro, Matteo Rossi, United nations global compact: Literature review and theory-based research agenda, J. Clean. Prod. 177 (2018) 633–654.

[17] Jingwei Ni, Julia Bingler, Chiara Colesanti-Senni, Mathias Kraus, Glen Gostlow, Tobias Schimanski, Dominik Stammbach, Saeid Ashraf Vaghefi, Qian Wang, Nicolas Webersinke, Tobias Wekhof, Tingyu Yu, Markus Leippold, Leippold chatreport: Democratizing sustainability disclosure analysis through llm-based tools, in: EMNLP 2023-2023 Conference on Empirical Methods in Natural Language Processing, Proceedings of the System Demonstrations, 2023, pp. 21–51.

[18] Gülçin Büyüközkan, Yağmur Karabulut, Sustainability performance evaluation: Literature review and future directions, J. Environ. Manag. 217 (2018) 253–267.

[19] Anne-Kathrin Hinze, Franziska Sump, Corporate social responsibility and financial analysts: A review of the literature, Sustain. Account. Manag. Policy J. 10 (1) (2019) 183–207.

[20] A. Søgaard, Explainable natural language processing, in: Synthesis Lectures on Human Language Technologies, Springer Cham, Cham, Switzerland, 2021.

[21] Doan Thi Thuc Nguyen, An empirical study on the impact of sustainability reporting on firm value, J. Competitiveness (2020).

[22] Indra Abeysekera, A framework for sustainability reporting, Sustain. Account. Manag. Policy J. 13 (6) (2022) 1386–1409.

[23] Marco Fasan, Annual reports, sustainability reports and integrated reports: Trends in corporate disclosure, in: Integrated Reporting: Concepts and Cases that Redefine Corporate Accountability, Springer, Cham, Switzerland, 2013, pp. 41–57.

[24] Gordon Kuo Siong Tan, Assembling sustainability reporting in singapore, Competition & Change 26 (5) (2022) 629–649.

[25] Tyler A. Scott, Nicholas Marantz, Nicola Ulibarri, Use of boilerplate language in regulatory documents: Evidence from environmental impact statements, J. Public Adm. Res. Theory 32 (2022) 576–590.

[26] Jeong-Bon Kim, Chong Wang, Feng Wu, The real effects of risk disclosures: Evidence from climate change reporting in 10-ks, Rev. Account. Stud.

[27] Jin Boon Wong, Qin Zhang, Stock market reactions to adverse esg disclosure via media channels, Br. Account. Rev. 54 (1) (2022) 101045.

[28] Adriana Barbeito-Caamaño, Ricardo Chalmeta, Using big data to evaluate corporate social responsibility and sustainable development practices, Corp. Soc. Responsib. Environ. Manag. 27 (6) (2020) 2831–2848.

[29] Felix Beske, Ellen Haustein, Peter C. Lorson, Materiality analysis in sustainability and integrated reports, Sustain. Account. Manag. Policy J. 11 (1) (2020) 162–186.

[30] Davide Fiaschi, Elisa Giuliani, Federica Nieri, Nicola Salvati, How bad is your company? Measuring corporate wrongdoing beyond the magic of esg metrics, Bus. Horiz. 63 (3) (2020) 287–299.

[31] Ivan Ruiz Manuel, Kornelis Blok, Quantitative evaluation of large corporate climate action initiatives shows mixed progress in their first half-decade, Nature Commun. 14 (1) (2023).

[32] Johannes Meuer, Julian Koelbel, Volker H. Hoffmann, On the nature of corporate sustainability, Organ. Environ. 33 (2020) 319–341.

[33] Felix Beske, Ellen Haustein, Peter C. Lorson, Materiality analysis in sustainability and integrated reports, Sustain. Account. Manag. Policy J. 11 (1) (2020) 162–186.

[34] Sebastião Vieira de Freitas Netto, Marcos Felipe Falcão Sobral, Ana Regina Bezerra Ribeiro, Gleibson Robert da Luz Soares, Concepts and forms of greenwashing: A systematic review, Environ. Sci. Eur. 32 (1) (2020) 1–12.

[35] Philippe Delacote, Tara L'Horty, Andreas Kontoleon, Thales A. West P., Anna Creti, Ben Filewod, Gwenole LeVelly, Alejandro Guizar-Coutiño, Ben Groom, Micah Elias, Strong transparency required for carbon credit mechanisms, Nature Sustainability (2024) 1–8.

[36] Eun-Hee Kim, Thomas Lyon, Greenwash vs. brownwash: Exaggeration and undue modesty in corporate sustainability disclosure, Organ. Sci. 26 (2015) 705–723.

[37] Kira R. Fabrizio, Eun-Hee Kim, Reluctant disclosure and transparency: Evidence from environmental disclosures, Organ. Sci. 30 (6) (2019) 1207–1231.

[38] Colin Higgins, Samuel Tang, Wendy Stubbs, On managing hypocrisy: The transparency of sustainability reports, J. Bus. Res. 114 (2020) 395–407.

[39] Ramona Zharfpeykan, Representative account or greenwashing? voluntary sustainability reports in Australia's mining/metals and financial services industries, Bus. Strategy Environ. 30 (4) (2021) 2209–2223.

[40] Nils Smeuninx, Bernard De Clerck, Walter Aerts, Measuring the readability of sustainability reports: A corpus-based analysis through standard formulae and nlp, Int. J. Bus. Commun. 57 (1) (2020) 52–85.

[41] Tushar Goel, Palak Jain, Ishan Verma, Lipika Dey, Shubham Paliwal, Mining company sustainability reports to aid financial decision-making, in: Proc. of AAAI Workshop on Know. Disc. from Unstructured Data in Fin. Services, 2020.

[42] Thomas A. Tsalis, Kyveli E. Malamateniou, Dimitrios Koulouriotis, Ioannis E. Nikolaou, New challenges for corporate sustainability reporting: United nations' 2030 agenda for sustainable development and the sustainable development goals, Corp. Soc. Responsib. Environ. Manag. 27 (4) (2020) 1617–1629.

[43] Ardian Qorri, Saranda Gashi, Andrzej Kraslawski, A practical method to measure sustainability performance of supply chains with incomplete information, J. Clean. Prod. 341 (2022) 130707.

[44] Erik Cambria, Understanding Natural Language Understanding, Springer, ISBN: 978-3-031-73973-6, 2024.

[45] Iti Chaturvedi, Yew-Soon Ong, Ivor Tsang, Roy Welsch, Erik Cambria, Learning word dependencies in text by means of a deep recurrent belief network, Knowledge-Based Systems 108 (2016) 144–154.

[46] Dominik Stammbach, Nicolas Webersinke, Julia Bingler, Mathias Kraus, Markus Leippold, Environmental claim detection, in: Anna Rogers, Jordan Boyd-Graber, Naoaki Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 1051–1066.

[47] Marissa Abram, Karen Mancini, R. Parker, Methods to integrate natural language processing into qualitative research, Int. J. Qual. Methods 19 (2020) 12.

[48] K.R. Chowdhary, Natural Language Processing, Springer India, New Delhi, India, 2020, pp. 603–649.

[49] Julia Anna Bingler, Mathias Kraus, Markus Leippold, Nicolas Webersinke, Cheap talk and cherry-picking: What climatebert has to say on corporate climate risk disclosures, Finance Res. Lett. 47 (2022) 102776.

[50] Marco Polignano, Nicola Bellantuono, Francesco Paolo Lagrasta, Sergio Caputo, Pierpaolo Pontrandolfo, Giovanni Semeraro, An NLP approach for the analysis of global reporting initiative indexes from corporate sustainability reports, in: Proceedings of the First Computing Social Responsibility Workshop Within the 13th Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 1–8.

[51] Jaeyoung Lee, Misuk Kim, Esg information extraction with cross-sectoral and multi-source adaptation based on domain-tuned language models, Expert Syst. Appl. 221 (2023) 119726.

[52] Juyeon Kang, Ismail El Maarouf, FinSim4-ESG shared task: Learning semantic similarities for the financial domain. extended edition to ESG insights, in: Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing, FinNLP, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 2022, pp. 211–217.

[53] Tobias Schimanski, Julia Bingler, Mathias Kraus, Camilla Hyslop, Markus Leippold, ClimateBERT-NetZero: Detecting and assessing net zero and reduction targets, in: Houda Bouamor, Juan Pino, Kalika Bali (Eds.), Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2023, pp. 15745–15756.

[54] Yusheng Zhou, Xueqin Wang, Kum Fai Yuen, Sustainability disclosure for container shipping: A text-mining approach, Transp. Policy 110 (2021) 465–477.

[55] Yosephine Susanto, Andrew Livingstone, Bee Chin Ng, Erik Cambria, The Hourglass Model revisited, IEEE Intelligent Systems 35 (5) (2020) 96–102.

[56] Stefan Pasch, Daniel Ehnes, Nlp for responsible finance: Fine-tuning transformer-based models for esg, in: 2022 IEEE International Conference on Big Data, Big Data, 2022, pp. 3532–3536.

[57] Yu-Min Tseng, Chung-Chi Chen, Hen-Hsen Huang, Hsin-Hsi Chen, Dynamicesg: A dataset for dynamically unearthing esg ratings from news articles, in: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, 2023, pp. 5412–5416.

[58] Jannik Fischbach, Max Adam, Victor Dzhagatspanyan, Daniel Mendez, Julian Frattini, Oleksandr Kosenkov, Parisa Elahidoost, Automatic esg assessment of companies by mining and evaluating media coverage data: Nlp approach and tool, 2022.

[59] Alik Sokolov, Jonathan Mostovoy, Jack Ding, Luis Seco, Building machine learning systems for automated esg scoring, J. Impact ESG Invest. 1 (3) (2021) 39–50.

[60] Min Gyeong Kim, Kyu Sung Kim, Kun Chang Lee, Analyzing the effects of topics underlying companies' financial disclosures about risk factors on prediction of esg risk ratings: Emphasis on bertopic, in: 2022 IEEE International Conference on Big Data, Big Data, 2022, pp. 4520–4527.

[61] Konstantin Ignatov, When esg talks: Esg tone of 10-k reports and its significance to stock markets, Int. Rev. Financ. Anal. 89 (2023) 102745.

[62] Jaime González-González, Silvia García-Méndez, Francisco De Arriba-Pérez, Francisco J. González-Castaño, Óscar Barba-Seara, Explainable automatic industrial carbon footprint estimation from bank transaction classification using natural language processing, IEEE Access 10 (2022) 126326–126338.

[63] Jinfang Tian, Qian Cheng, Rui Xue, Yilong Han, Yuli Shan, A dataset on corporate sustainability disclosure, Sci. Data 10 (1) (2023) 182.

[64] Darío Garigliotti, Sdg target detection in environmental reports using retrieval-augmented generation with llms, in: Proceedings of the 1st Workshop on Natural Language Processing Meets Climate Change, ClimateNLP 2024, 2024, pp. 241–250.

[65] Elisabeth Sinnewe, Troy Yao, Mahbub Zaman, Informing or obfuscating stakeholders: Integrated reporting and the information environment, Bus. Strategy Environ. 30 (8) (2021) 3893–3906.

[66] Dan Bi, Ju-e Guo, Erlong Zhao, Shaolong Sun, Shouyang Wang, Using word embedding for environmental violation analysis: Evidence from pennsylvania unconventional oil and gas compliance reports, Environ. Dev. 47 (2023) 100905.

[67] Peter M. Clarkson, Jordan Ponn, Gordon D. Richardson, Frank Rudzicz, Albert Tsang, Jingjing Wang, A textual analysis of us corporate social responsibility reports, Abacus 56 (1) (2020) 3–34.

[68] Srishti Mehra, Robert Louka, Yixun Zhang, Esgbert: Language model to help with classification tasks related to companies environmental, social, and governance practices, 2022, arXiv preprint arXiv:2203.16788.

[69] Rui Mao, Kai He, Xulang Zhang, Guanyi Chen, Jinjie Ni, Zonglin Yang, Erik Cambria, A survey on semantic processing techniques, Inf. Fusion 101 (2024) 101988.

[70] Jeffrey Pennington, Richard Socher, Christopher D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2014, pp. 1532–1543.

[71] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.

[72] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al., Gpt-4 technical report, 2023, arXiv preprint arXiv:2303.08774.

[73] Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, Hsiao-Wuen Hon, Unified language model pre-training for natural language understanding and generation, Adv. Neural Inf. Process. Syst. 32 (2019).

[74] Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven.Le Scao, Arun Raja, et al., Multitask prompted training enables zero-shot task generalization, in: International Conference on Machine Learning, 2022.

[75] Rui Mao, Guanyi Chen, Xulang Zhang, Frank Guerin, Erik Cambria, GPTEval: A survey on assessments of chatGPT and GPT-4, in: Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, Nianwen Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC-COLING 2024, ELRA and ICCL, Torino, Italia, 2024, pp. 7844–7866.

[76] Xulang Zhang, Rui Mao, Erik Cambria, A survey on syntactic processing techniques, Artif. Intell. Rev. 56 (2023) 5645–5728.

[77] Alebachew Chiche, Betselot Yitagesu, Part of speech tagging: A systematic review of deep learning and machine learning approaches, J. Big Data 9 (1) (2022) 10.

[78] Zuchao Li, Hai Zhao, Kevin Parnow, Global greedy dependency parsing, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 8319–8326.

[79] Kaiyu Yang, Jia. Deng, Strongly incremental constituency parsing with graph neural networks, Adv. Neural Inf. Process. Syst. 33 (2020) 21687–21698.

[80] Rui Mao, Xiao Li, Bridging towers of multi-task learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, 35, 2021, pp. 13534–13542.

[81] Sergio Consoli, Luca Barbaglia, Sebastiano Manzan, Fine-grained, aspect-based sentiment analysis on economic and financial lexicon, Knowl.-Based Syst. 247 (2022) 108781.

[82] Adrien Ehrhardt, Minh Tuan Nguyen, Automated esg report analysis by joint entity and relation extraction, in: Michael Kamp, Irena Koprinska, Adrien Bibal, Tassadit Bouadi, Benoît Frénay, Luis Galárraga, José Oramas, Linara Adilova, Yamuna Krishnamurthy, Bo Kang, Christine Largeron, Jefrey Lijffijt, Tiphaine Viard, Pascal Welke, Massimiliano Ruocco, Erlend Aune, Claudio Gallicchio, Gregor Schiele, Franz Pernkopf, Michaela Blott, Holger Fröning, Günther Schindler, Riccardo Guidotti, Anna Monreale, Salvatore Rinzivillo, Przemyslaw Biecek, Eirini Ntoutsi, Mykola Pechenizkiy, Bodo Rosenhahn, Christopher Buckley, Daniela Cialfi, Pablo Lanillos, Maxwell Ramstead, Tim Verbelen, Pedro M. Ferreira, Giuseppina Andresini, Donato Malerba, Ibéria Medeiros, Philippe Fournier-Viger, M. Saqib Nawaz, Sebastian Ventura, Meng Sun, Min Zhou, Valerio Bitetta, Ilaria Bordino, Andrea Ferretti, Francesco Gullo, Giovanni Ponti, Lorenzo Severini, Rita Ribeiro, João Gama, Ricard Gavaldà, Lee Cooper, Naghmeh Ghazaleh, Jonas Richiardi, Damian Roqueiro, Diego Saldana Miranda, Konstantinos Sechidis, Guilherme Graça (Eds.), Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Springer International Publishing, Cham, 2021, pp. 325–340.

[83] Kelvin Du, Frank Xing, Rui Mao, Erik Cambria, Finsenticnet: A concept-level lexicon for financial sentiment analysis, in: 2023 IEEE Symposium Series on Computational Intelligence, SSCI, IEEE, 2023, pp. 109–114.

[84] E. Cambria, R. Mao, M. Chen, Z. Wang, S. Ho, Seven pillars for the future of artificial intelligence, IEEE Intell. Syst. 38 (06) (2023) 62–69.

[85] Hendrik Schuff, Alon Jacovi, Heike Adel, Yoav Goldberg, Ngoc.Thang. Vu, Human interpretation of saliency-based explanation over text, in: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022, pp. 611–636.

[86] Wei Jie Yeo, Ranjan Satapathy, Rick Siow Mong Goh, Erik Cambria, How interpretable are reasoning explanations from prompting large language models? in: NAACL Findings, pp. 2148–2164.

[87] Yibing Liu, Haoliang Li, Yangyang Guo, Chenqi Kong, Jing Li, Shiqi Wang, Rethinking attention-model explainability through faithfulness violation test, in: International Conference on Machine Learning, PMLR, 2022, pp. 13807–13824.

[88] Hugues Turbé, Mina Bjelogrlic, Christian Lovis, Gianmarco. Mengaldo, Evaluation of post-hoc interpretability methods in time-series classification, Nat. Mach. Intell. 5 (3) (2023) 250–260.

[89] Jiawen Wei, Hugues Turbé', Gianmarco Mengaldo, Revisiting the robustness of post-hoc interpretability methods, 2024, arXiv preprint arXiv:2407.19683.

[90] Sara Hooker, Dumitru Erhan, Pieter-Jan Kindermans, Been Kim, A benchmark for interpretability methods in deep neural networks, Adv. Neural Inf. Process. Syst. 32 (2019).

[91] Zachary C. Lipton, The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery, Queue 16 (3) (2018) 31–57.

[92] Gianmarco Mengaldo, Explain the black box for the sake of science: Revisiting the scientific method in the era of generative artificial intelligence, 2024, arXiv preprint arXiv:2406.10557.

[93] Hugues Turbé, Mina Bjelogrlic, Gianmarco Mengaldo, Christian. Lovis, Protos-vit: Visual foundation models for sparse self-explainable classifications, 2024, arXiv preprint arXiv:2406.10025.

[94] Sooji Han, Rui Mao, Erik Cambria, Hierarchical attention network for explainable depression detection on Twitter aided by metaphor concept mappings, in: Proceedings of the 29th International Conference on Computational Linguistics, COLING, International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 94–104.

[95] Mireia Guix, Claudia Ollé, Xavier Font, Trustworthy or misleading communication of voluntary carbon offsets in the aviation industry, Tour. Manag. 88 (2022) 104430.

[96] Xulang Zhang, Rui Mao, Kai He, Erik Cambria, Neurosymbolic sentiment analysis with dynamic word sense disambiguation, in: Findings of the Association for Computational Linguistics: EMNLP 2023, 2023, pp. 8772–8783.

[97] Charles H. Cho, Robin W. Roberts, Dennis M. Patten, The language of us corporate environmental disclosure, Account. Organ. Soc. 35 (2010) 431–443.

[98] Jeong-Bon Kim, Chong Wang, Feng Wu, The real effects of risk disclosures: Evidence from climate change reporting in 10-ks, Rev. Account. Stud. 28 (4) (2023) 2271–2318.

[99] Peter Nickl, Lu Xu, Dharmesh Tailor, Thomas Möllenhoff, Mohammad Emtiyaz E. Khan, The memory-perturbation equation: Understanding model' s sensitivity to data, in: Advances in Neural Information Processing Systems, vol. 36, Curran Associates, Inc, New York, 2023, pp. 26923–26949.

[100] Manish Raghavan, Solon Barocas, Jon Kleinberg, Karen Levy, Mitigating bias in algorithmic hiring: Evaluating claims and practices, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, pp. 469–481.