Detecting Greenwashing in the Environmental, Social, and Governance Domains Using Natural Language Processing

Yue Zhao[1], Leon Kroher[2], Maximilian Engler[2] and Klemens Schnattinger[2]
[1]*Intergration Alpha GmbH, Fabrikstrasse 5, 6330 Cham, Switzerland*
[2]*Business Innovation Center, Baden-Wuerttemberg Cooperative State University (DHBW),*
Hangstraße 46-50, 79539 Loerrach, Germany

Keywords: Greenwashing, Natural Language Processing (NLP), Environmental, Social, and Governance (ESG), Sentiment Analysis, Question-and-Answer Generation, Pharmaceutical Firms, Public Perception, Social Media, Monitoring Mechanisms.

Abstract: Greenwashing, where companies misleadingly project environmental, social, and governance (ESG) virtues, challenges stakeholders. This study examined the link between internal ESG sentiments and public opinion on social media across 12 pharmaceutical firms from 2012 to 2022. Using natural language processing (NLP), we analyzed internal documents and social media. Our findings showed no significant correlation between internal and external sentiment scores, suggesting potential greenwashing if there's inconsistency in sentiment. This inconsistency can be a red flag for stakeholders like investors and regulators. In response, we propose an NLP-based Q&A system that generates context-specific questions about a company's ESG performance, offering a potential solution to detect greenwashing. Future research should extend to other industries and additional data sources like financial disclosures.

1 INTRODUCTION

1.1 Research Objective and Hypothesis

This study aims to probe the capability of diverse Natural Language Processing (NLP) frameworks in pinpointing greenwashing activities within the Environmental, Social, and Governance (ESG) sphere (Kim & Lyon, 2015). Greenwashing refers to the deceptive portrayal of a firm's performance in environmental, social, or governance facets. Accurately detecting such actions is pivotal for stakeholders like investors and regulators, ensuring they gauge a company's genuine dedication to sustainability (Delmas & Burbano, 2011). To accomplish this, we evaluate 12 pharmaceutical entities based on their 2021 revenue according to Fortune (2021). Additionally, we introduce innovative mechanisms tailored for automated greenwashing surveillance.

Our thesis suggests that a diminished correlation between sentiment metrics from in-house corporate resources and external social media narratives hints at discord between internal strategic utterances and collective public sentiment (Lyon & Montgomery, 2015). By utilizing the FinBERT-ESG-9-Categories model pioneered by Huang et al., we aspire to shed light on potential incongruences in ESG narratives, hinting at latent greenwashing (Huang et al., 2022).

Through this analytical lens, our research aspires to shed light on the proficiency of NLP frameworks in discerning greenwashing undertakings in the ESG realm, and in the evolution of automated surveillance systems for a nuanced sustainability performance appraisal (Asif et al., 2023).

1.2 Greenwashing Detection in ESG via NLP

Utilizing NLP methodologies for greenwashing detection in the ESG sphere entails employing sophisticated computational approaches to scrutinize and decipher textual data associated with a corporation's sustainability endeavours (Davenport & Harris, 2019). By capitalizing on NLP techniques, investigators and interested parties can unveil potential inconsistencies between a firm's internal strategic communications and public perceptions, potentially signifying attempts to overstate or

Zhao, Y., Kroher, L., Engler, M. and Schnattinger, K.
Detecting Greenwashing in the Environmental, Social, and Governance Domains Using Natural Language Processing.
DOI: 10.5220/0012155400003598
In *Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2023) - Volume 1: KDIR*, pages 175-181
ISBN: 978-989-758-671-2; ISSN: 2184-3228

misrepresent ESG accomplishments (Orlitzky, 2011). Greenwashing detection through NLP may encompass techniques such as sentiment analysis, topic modeling, text classification, and knowledge graph (Liu, 2012). These approaches facilitate the extraction of valuable insights from diverse data sources, including corporate reports, press releases, and social media content (Siano et al., 2021). By integrating NLP methods in the examination of ESG-centric data, researchers can devise more rigorous and dependable mechanisms for identifying greenwashing practices, ultimately fostering enhanced transparency and accountability in the corporate sustainability domain (Delmas & Burbano, 2011).

2 LITERATURE REVIEW

2.1 Greenwashing

Researchers have extensively studied methods to identify and quantify greenwashing, specifically focusing on the discrepancies between a company's public statements and its actual ESG actions (Marquis et al., 2016). Recent efforts have pivoted towards exploiting NLP techniques for greenwashing identification within the ESG framework. Specifically, Moodaley et al. (2023) employed bibliometric and thematic analysis to probe the nexus between greenwashing, sustainability reporting, and the confluence of AI and ML in scholarly works. Woloszyn et al. (2021) highlighted the dynamic interplay between human-driven and machine-driven computing in detecting green claims, accentuating the rising significance of machines in the process. The burgeoning field of automated tools designed to detect greenwashing is reshaping the research landscape. These cutting-edge systems facilitate a continuous analysis of ESG narratives, providing stakeholders with a more immediate and efficient means to spot potential greenwashing activities (Starik et al., 2016). Nugent et al. (2020) explored the use of pre-trained and fine-tuned models to categorize ESG topics. Despite the initial strides in harnessing NLP to detect ESG-focused greenwashing, there remains a void in the literature. Specifically, there's a noticeable dearth of studies that target the accurate identification of greenwashing and the enhancement of automated monitoring tools, especially within the pharmaceutical sector, through NLP approaches. Building upon this, our research delves deeper by not only examining ESG topics but also by introducing sentiment analysis and anticipatory monitoring techniques as well as endeavors to fill this research lacuna.

2.2 FinBERT

FinBERT, short for Financial BERT, is a specialized language model based on the BERT (Bidirectional Encoder Representations from Transformers) architecture, tailored for financial sentiment analysis. It is pre-trained on a large corpus of financial text data to capture the nuances and context found in financial documents and reports. FinBERT allows for the extraction of sentiment information from financial texts, which can be useful in various financial applications such as predicting stock prices, evaluating corporate performance, and detecting potential greenwashing practices (Huang et al., 2022).

3 METHODOLOGY AND TOOLS

3.1 Methodology

Our research framework delves deep into the use of Natural Language Processing (NLP) for the detection and monitoring of ESG-centric greenwashing activities. We sourced textual data from internal corporate disclosures and Twitter, gaining a dual-pronged perspective. After data extraction, a thorough preprocessing phase was initiated, incorporating tokenization, stopwords removal, and lemmatization, to ready the data for subsequent analysis. We then utilized the FinBERT-ESG-9-Categories model to categorize the data. Sentiment analysis was performed using TextBlob. A correlation coefficient analysis was then conducted to identify disparities indicative of greenwashing. Further, we introduced an innovative NLP-driven framework for ESG greenwashing monitoring, transforming core ESG data into structured question-and-answer pairs. This comprehensive approach utilized tools such as spaCy, AllenNLP, and NLTK, ensuring methodological robustness. Our methodology's ultimate goal is to enhance transparency and accountability in the ESG domain.

3.2 NLP Tools

3.2.1 spaCy

spaCy stands as an avant-garde NLP library known for its stellar performance. Designed for professional-grade applications, its attributes of rapid processing, user-friendliness, and operational efficiency have made it a mainstay among the developer and research fraternities. Detailed insights and documentation can be procured from https://spacy.io.

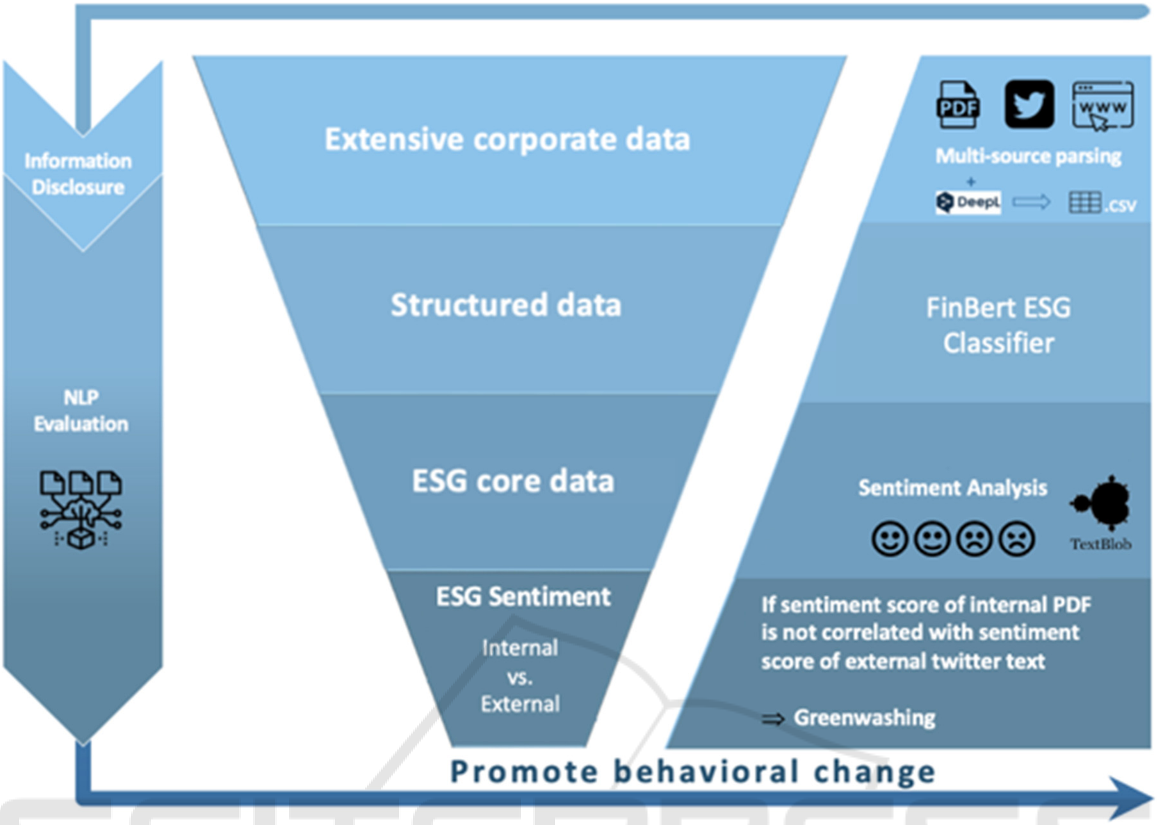

Figure 1: NLP framework to identify ESG-Greenwashing.

3.2.2 AllenNLP

A product of the Allen Institute for Artificial Intelligence, AllenNLP is a distinguished library underpinned by PyTorch. It is crafted with a focus on facilitating deep learning endeavors within the NLP domain. The library offers superior abstractions and tools tailored for the construction of sophisticated NLP models. Comprehensive details can be explored at https://allenai.org/allennlp.

3.2.3 NLTK

The Natural Language Toolkit (NLTK) serves as an expansive platform supporting the development of Python-driven applications for linguistic data analysis. With interfaces to an extensive range of corpora and lexical databases like WordNet, NLTK also offers an arsenal of text manipulation tools. These tools cover a spectrum of tasks, ranging from tokenization and stemming to semantic reasoning. For an exhaustive list of features and utilities, one can refer to https://www.nltk.org/.

4 NLP FRAMEWORK TO IDENTIFY ESG-GREENWASHING

The workflow of the NLP Framework to identify ESG-Greenwashing is shown in Figure 1. Each step will be described in detail below.

4.2 Data Collection and Preparation

Detecting greenwashing tactics within the ESG landscape via NLP methodologies necessitates sourcing and optimizing relevant textual data from diverse channels. A vital first step, this guarantees the data's suitability for in-depth analysis. Primary data reservoirs encompass internal corporate disclosures (Agyei-Mensah, 2016) and external insights derived from Twitter (Goodell et al., 2019). Together, they present a holistic understanding of an enterprise's ESG undertakings, their efficacy, and any emergent conflicts (Moodaley & Telukdarie, 2023). As illustrated in Figure 1, official corporate websites typically host publicly accessible internal documents in their sustainability or ESG sections, usually in PDF format. Additionally, Twitter content is harvested using web scraping tools via APIs (Goodell et al., 2019). Any non-English data extracted are subsequently translated into English using DeepL, an online translation service that uses artificial intelligence and neural networks to provide high-quality translation between multiple languages, which renders them primed for analysis (DeepL, 2023). The data underwent essential preprocessing stages prior to analysis. These stages included tokenization, the removal of stopwords, lemmatization or stemming, the exclusion of special characters and numbers, and converting text to ensure uniformity across the dataset (Gautam et al., 2022). From this process, we extracted a total of 65,763 data points from PDF documents and 118,560 data points from Twitter. In the PDF dataset, each data point corresponds to an individual sentence, whereas in the Twitter dataset, each data point represents a full tweet.

4.3 Data Analysis

In our quest to thoroughly analyze the primary ESG Data, we followed a systematic methodology.

Firstly, leveraging the specialized capabilities of the FinBERT-ESG-9-Categories model, we meticulously classified our dataset into relevant ESG themes. These include Climate Change, Natural Capital, Pollution & Waste, Human Capital, Product Liability, Community Relations, Corporate Governance, Business Ethics & Values, and a catch-all non-ESG category, as expounded by Huang et al. (2022).

With our data now suitably categorized, we proceeded to evaluate its emotional undertones using TextBlob, a lexicon-focused sentiment analysis library. It assigns sentiment scores to textual content, ranging from negative to positive (Liu, 2012). The essence of sentiment analysis lies in its ability to measure the affective orientation of opinions embedded within texts (Liu and Zhang, 2012). By applying TextBlob to our segmented data, we could gauge the general mood encompassing corporate internal ESG declarations and juxtapose this with the prevailing sentiment on social platforms like Twitter (Pang and Lee, 2008).

Each of the ESG themes, from Climate Change to Business Ethics & Values, underwent in-depth sentiment scrutiny. This allowed us to extract correlations (or the lack thereof) between the sentiments expressed in internal corporate reports and those voiced by the public on Twitter.

To move beyond a simple sentiment score and to deeply understand the dynamics between internal corporate sentiments and external public opinions, we employed correlation coefficient analysis. This method, highlighted in research by Yu et al. (2023), offers insights into the strength and direction of the relationship between these two sentiment sources. By adopting this approach, we were able to pinpoint discrepancies that might hint at greenwashing practices, a phenomenon outlined by Delmas and Burbano (2011).

Through this layered analytical methodology, our study aimed to shed light on potential divergences between corporate ESG narratives and public perceptions, offering stakeholders a clearer picture of potential greenwashing instances.

4.4 Research Findings

Our findings reveal a discernible disconnect between the sentiments expressed in internal ESG strategies and those voiced in public opinions on social media. The internal sentiment scores, derived from official PDF documents, exhibit a predominantly positive and stable trajectory, contrasting with the more volatile sentiment evident in Twitter scores. By juxtaposing the sentiment scores from official documents with those on Twitter, we computed correlation coefficients. For example, the "pollution and waste" category exemplifies this discrepancy most prominently, as depicted in Figure 2. Over the period from 2012 to 2022, most companies showcased in Figure 2 have a negative linear correlation between their official sentiment scores and Twitter. While the general trend indicates a divergence, a notable exception is Company 4. Between 2012 and 2014, its official sentiment scores and Twitter scores exhibited a mild positive correlation. Cumulatively, these results underscore a discord between a company's internal ESG declarations and the broader public perception, underscoring the urgency for deeper scrutiny of potential greenwashing activities.

5 NLP FRAMEWORK TO MONITOR ESG-GREENWASHING

To address the burgeoning need for vigilant greenwashing monitoring, we present a framework driven by NLP, specifically designed for ESG greenwashing detection and assessment. This framework harnesses the essence of core ESG data extracted from internal corporate documents and adeptly transforms this data into organized

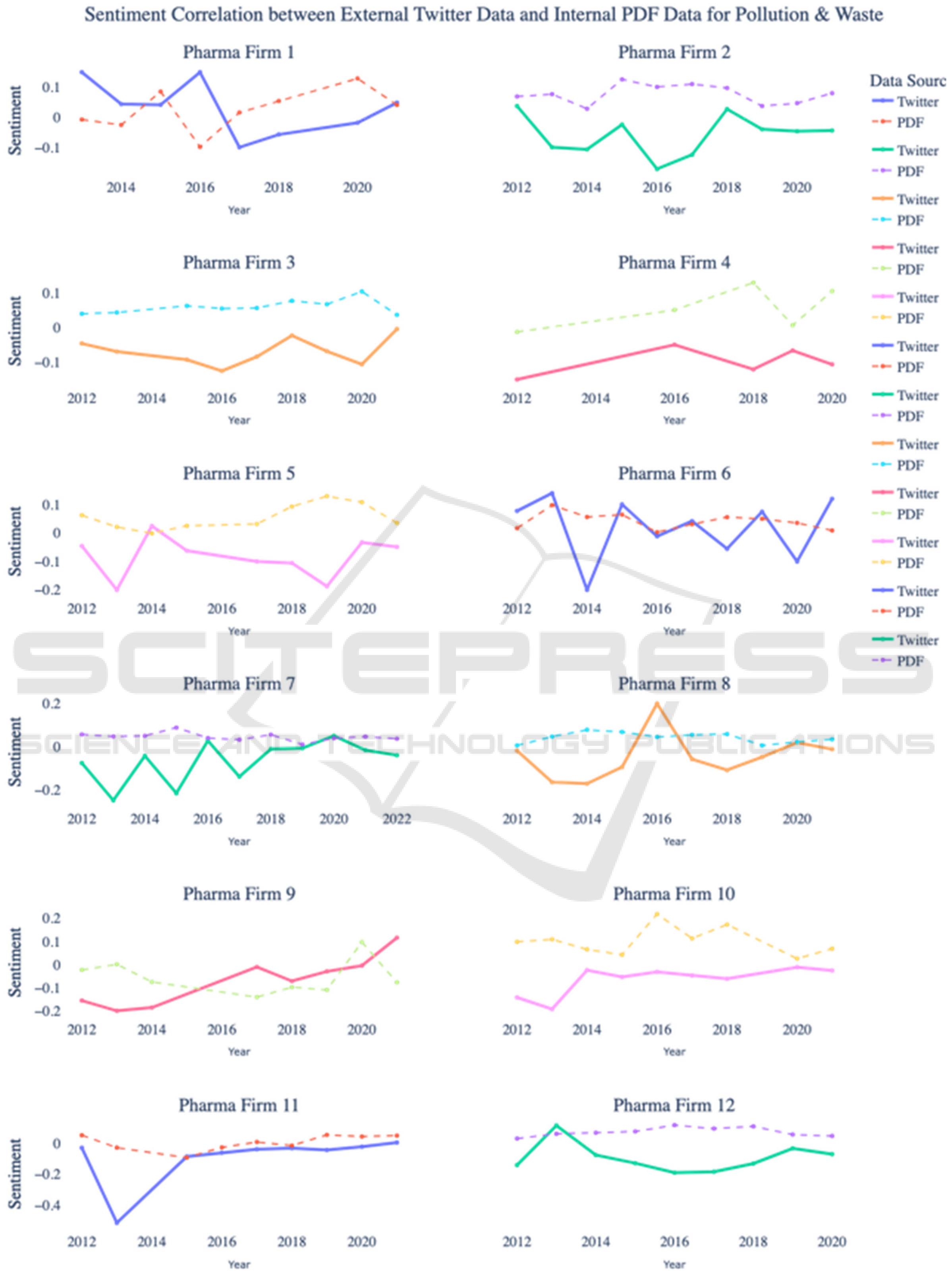

Figure 2: Sentiment correlation between internal PDF data and external Twitter data for pollution & waste.

question-and-answer pairs. This structure immensely simplifies the process of internally validating ESG claims. When enterprises interact with our system by posing queries, it efficiently correlates these queries with our pre-established database, a product of our data collection. As a result, it swiftly prese pertinent, pre-formulated questions coupled with their corresponding answers, offering a seamless mechanism to verify and substantiate ESG assertions.

To further refine the raw ESG core data, we incorporate prominent NLP tools including spaCy, AllenNLP, and NLTK. Given that each entry in the core data equates to an entire paragraph, the first phase entails paragraph parsing. This involves segmenting the paragraph, resolving coreferences and abbreviations, and determining textual entailment—tasks adeptly handled by AllenNLP. One might question the handling of ambiguities or contradictory information during this phase. Our system has been trained on a vast corpus of ESG data, enabling it to identify and flag such ambiguities for human review. In the event of potential mismatches or contradictions within the text, the system triggers an internal review protocol where a human expert can intervene to ensure accuracy.

The ensuing phase involves sentence-level parsing. This includes sentence segmentation, punctuation realignment, constituency parsing, and semantic role labeling, tasks for which spaCy is particularly well-suited. Each sentence, in essence, is dissected into its granular components, providing the foundational bricks upon which our subsequent question-and-answer pairs are built. Post these parsing procedures, the crux of our methodology surfaces: the generation of question-and-answer pairs. These pairs are formulated by gleaning the pivotal ESG-related assertions within the internal corporate data. For illustration, if an enterprise's internal dossier mentions a decrease in carbon footprints, our system could potentially generate a pair such as:

QUESTION: Over the decade from 2009 to 2019, by what percentage have we curtailed our carbon emissions?

ANSWER: 50% reduction

It's important to underline how answers are crafted. They aren't merely extracted in a raw form and serve as repositories of information but are a product of the system's deep semantic understanding of the input data, they act as a litmus test for veracity. The framework intelligently discerns context, evaluates claims against benchmarks, and composes answers that are succinct and accurate. This question-and-answer infrastructure not only acts as a self-monitoring tool but also as a real-time verifier, returning vetted answers from the database in response to stakeholders' inquiries, hence promoting transparency and accountability.

6 CONCLUSIONS

This study ventures into the untapped potential of NLP methodologies in pinpointing potential greenwashing tendencies within the ESG landscape, an area of immense significance to stakeholders' intent on discerning the authenticity of corporate sustainability pledges. Our analysis draws parallels between the sentiment scores from internal ESG strategies and the public sentiment aired on Twitter for 12 pharmaceutical giants spanning a decade (2012-2022). Stemming from this analysis, we designed an innovative NLP-driven question-and-answer system, envisioned to expedite and enhance the process of monitoring greenwashing.

The findings illuminate a palpable disjunction between internal ESG sentiments and those manifested on Twitter. While internal sentiments predominantly radiate positive vibes and maintain consistency, their external counterparts on Twitter exhibit more fluctuations. The pollution and waste sector showcases a pronounced incongruity in sentiment alignment. This disparity signals a potential disconnect between corporate ESG proclamations and the prevailing public sentiment, accentuating the necessity for meticulous scrutiny of potential greenwashing practices. Our pioneering NLP blueprint, capitalizing on core ESG data to forge pertinent question-and-answer pairings, emerges as a formidable tool in the arsenal against greenwashing.

However, this research is not without its caveats. Primarily, our lens is confined to the pharmaceutical arena, analyzing a sample of merely 12 companies. The vistas for future inquiries are expansive, ranging from diversifying the industries under study to assimilating an eclectic mix of data sources like financial disclosures and third-party ESG evaluations. Furthermore, there lies a rich tapestry of sentiment analysis tools—like Vader, NLTK, and more—that beckon a deeper dive, perhaps integrating advanced deep learning architectures to bolster detection precision. The NLP model we've rolled out serves as a groundwork, ripe for tailoring to cater to sector-specific requisites. Additionally, the inaccessible nature of certain internal corporate datasets curtails our ability to present a more robust statistical quantitative analysis, leading our findings to emphasize potential greenwashing red flags rather than unequivocal greenwashing occurrences. A fascinating trajectory for future probes might involve juxtaposing our system’s outputs with results gleaned from avant-garde platforms like OpenAI. In summation, while this work establishes the plausibility of greenwashing detection, any concrete identification and subsequent mitigation strategies necessitate the proactive engagement of the corporate entities in question.

REFERENCES

Agyei-Mensah, B. K. (2016). Internal control information disclosure and corporate governance: evidence from an emerging market. *Corporate Governance: The international journal of business in society*, 16(1), 79-95.

Asif, M., Searcy, C., & Castka, P. (2023). ESG and Industry 5.0: The role of technologies in enhancing ESG disclosure. Technological Forecasting and Social Change, 195, 122806.

Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, 54(1), 64-87.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv*:1810.04805.

Davenport, T. H., & Harris, J. G. (2019). Competing on analytics: Updated, with a new introduction: The new science of winning. *Harvard Business Review Press.*

DeepL. (2023). Why DeepL Pro? *DeepL.* https://www.deepl.com/en/why-deepl-pro

Fortune. (2021). Fortune 500. https://fortune.com/ranking/fortune500/

Goodell, G., & Aste, T. (2019). A decentralized digital identity architecture. *Frontiers in Blockchain*, 2, 17.

Gautam, A. K., & Bansal, A. (2022). Performance analysis of supervised machine learning techniques for cyberstalking detection in social media. Journal of *Theoretical and Applied Information Technology*, 100(2), 449-461.

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. Contemporary Accounting Research, 40(2), 806-841.

Kim, E. H., & Lyon, T. P. (2015). Greenwash vs. brownwash: Exaggeration and undue modesty in corporate sustainability disclosure. Organization Science, 26(3), 705-723.

Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. *Organization & Environment*, 28(2), 223-249.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. *In Mining text data* (pp. 415-463). Springer, Boston, MA.

Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, norms, and selective disclosure: A global study of greenwashing. *Organization Science*, 27(2), 483-504.

Moodaley, W., & Telukdarie, A. (2023). Greenwashing, Sustainability Reporting, and Artificial Intelligence: A Systematic Literature Review. *Sustainability*, 15(2), 1481.

Nugent, T., Stelea, N., & Leidner, J. L. (2020). Detecting ESG topics using domain−specific language models and data augmentation approaches. *arXiv preprint arXiv*:2010.08319.

Orlitzky, M. (2011). Institutional logics in the study of organizations: The social construction of the relationship between corporate social and financial performance. Business Ethics Quarterly, 21(3), 409-444.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieva*l, 2(1–2), 1-135.

Schnattinger, K., Walterscheid, H. (2017). Opinion Mining Meets Decision Making: Towards Opinion Engineering. *In Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - Volume 1: KDIR*, pages 334-341.

Siano, A., Vollero, A., Conte, F., & Amabile, S. (2021). “More than words”: Expanding the taxonomy of greenwashing after the Volkswagen scandal. *Journal of Business Research*, 117, 577-586.

Starik, M., Kanashiro, P., & Collins, E. (2016). Sustainability management textbooks: Potentials, limitations, and future directions. *Organization & Environment*, 29(1), 69-95.

Woloszyn, V., Kobti, J., & Schmitt, V. (2021). Towards Automatic Green Claim Detection. *In Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation*, 28−34.

Yu, H., Liang, C., Liu, Z., & Wang, H. (2023). News-based ESG sentiment and stock price crash risk. *International Review of Financial Analysis*, 88, 102646.