Delft University of Technology

&

Leiden University

Industrial Ecology

Master's Thesis

Promises vs. Performance

A Multi-Dimensional Greenwashing Risk Assessment Tool Integrating Environmental Performance Data with NLP Communication Analysis

Author:
H.A. Baartman
Student ID: 4666429 (Delft))
Student ID: 3652882 (Leiden

Supervisors:
Dr. T. Chatzivasileiadis
Faculty of Technology, Policy and Management
Delft University of Technology

Dr.ir. A. Ghorbani
Faculty of Technology, Policy and Management
Delft University of Technology

August 18, 2025

Digital Resources

An electronic version of this thesis can be found at:
https://repository.tudelft.nl/

The related code modules and implementation guide can be found at:
https://github.com/AadB010/Greenwashing-Risk-Assessment-Tool

This repository contains the complete software implementation of the Greenwashing Risk Assessment Tool, including modular code components and detailed documentation for practical application.

Executive Summary

Clean energy spending needs to reach $4.5 trillion per year by 2030 to maintain a realistic chance of limiting global warming, yet current investment represents just 37% of what is needed for net zero by 2050. This and other energy transition requirements put enormous pressure on electric utilities to transform their business models while maintaining energy security and attracting investment, creating conditions ripe for potential greenwashing as companies might oversell environmental progress or engage in misleading communication strategies to stay competitive. Three fundamental gaps have limited systematic greenwashing detection: the separation between performance and communication analysis, single-dimensional communication measurement, and the lack of sector-specific methods.

This research addresses the question: *What greenwashing risk assessment tool can be developed for European electric utility companies based on environmental performance metrics and multi-dimensional communication analysis?* This Master thesis research project developed and tested the first systematic Greenwashing Risk Assessment Tool (GRAT), designed specifically for electric utilities but adaptable to other sectors through modified inputs, weights, and terminology. The GRAT bridges theoretical gaps by integrating quantitative performance measurement with multi-dimensional communication analysis. Constrained ensemble methodology is used to handle weight uncertainty for both performance scoring and greenwashing risk. The greenwashing risk is measured by five different dimensions: performance-communication gap, substantiation weakness, language vagueness, temporal orientation, and reporting consistency.

The GRAT follows a clear three-step process that can be understood and adapted without advanced programming skills. Users gather environmental data from several sources: third-party verified information like Refinitiv Eikon, self-reported metrics like CDP, and English sustainability reports. Once this data is collected, the tool automatically handles performance scoring, runs communication analysis through rule-based NLP methods, and produces integrated risk assessments. Rather than using completely arbitrary weights, the constrained ensemble approach tests thousands of valid weight combinations within available theoretical boundaries. Users can adjust weights based on their specific priorities or new research findings, since some component weightings still rely on limited literature guidance. When applying the tool to different sectors, users can adjust performance metrics to fit industry standards, update communication terminology lists, and modify component weights based on what matters most in that sector. The tool offers two validation approaches: internal validation uses sensitivity analysis to check how robust results are across different weight combinations, while external validation tests accuracy against real documented cases.

The tool was tested on 14 European electric utility companies during 2021-2022, a period influenced by COVID-19 disruptions and pre-CSRD regulations that should be considered when interpreting findings. The companies were selected based on data availability rather than sector representativeness, which means results reflect patterns among companies with comprehensive data rather than the broader utility sector. Among these 14 companies, the analysis found significant environmental performance differences. Scores ranged from 15.0 to 95.0 points, with year-over-year changes spanning from -35.1 to +63.0 points. Companies used different communication strategies that did not align with their performance scores. Only 36% showed temporal changes where performance and communication moved in the same direction, revealing a pattern where communication strategies often work independently from actual environmental achievements. When combining both aspects including the other communication quality dimensions, risk scores fell between 16.5 and 81.3. With no companies reaching the theoretical boundaries of 0 or 100, this suggests the dimensions capture different aspects

of greenwashing behaviour rather than overlapping measures.

The validation approach is methodologically sound, though statistical limitations arise from the small sample size. External validation provides indicative evidence that the GRAT may distinguish between companies with documented greenwashing accusations and those with clean records, but Mann-Whitney U testing with 14 observations has low statistical power and results could represent random phenomena between 2021-2022. Future validation with larger sample sizes would be required to establish statistically robust evidence of the tool's discriminant validity. The GRAT enables regulators to screen companies requiring investigation rather than providing conclusive evidence of greenwashing. This helps investors assess risk patterns in sustainability portfolios and offers researchers a tool adaptable to other sectors. The rule-based approach provides transparency suited for regulatory contexts while the constrained ensemble methodology addresses weight uncertainty problems in composite indicator development.

Acknowledgements

At the beginning of this thesis and during the thesis preparation course, I was genuinely nervous about finding a good research topic and creating work I could be proud of. The initial uncertainty felt a bit heavy, and I wondered whether I could produce something meaningful. As the process went on, after struggling with dataset limitations and having to exclude many companies from my sample due to data constraints, a clearer picture started to form one piece at a time. What began as uncertainty became a thesis I genuinely enjoyed writing and refining. I discovered that this way of thinking, while hopefully making a difference for a better future, was something I was good at and truly liked spending time on.

I would like to thank both my supervisors for their guidance throughout this process. Their openness and friendliness made every meeting something I looked forward to, and their willingness to think along with me helped shape this research in ways I could not have managed alone.

Theo, my first supervisor, thank you for meeting with me so often and for being incredibly flexible when I needed to move meetings at the last minute. Your availability during holidays to answer questions and give feedback went far beyond what I expected. Your tips on data sources and management were invaluable in navigating the complex world of sustainability data.

Amineh, thank you for providing the opportunity to do this research in the first place. Granting access to certain parts of the CDP database was where it all started. Your feedback on methodological changes and research structure, along with the very specific suggestions you provided, helped me greatly in developing both the analytical approach and the overall direction.

I would like to thank Leiden University and TU Delft for this master's program. It was even more interesting than I expected, and I enjoyed every part of it. My gratitude also goes to contacts at Erasmus University Rotterdam for granting me access to the Datastream platform for the third-party verification component of this research. This turned out to be the most influential pivot point during the whole process.

Finally, I want to thank my parents and my girlfriend Daantje for their support during this thesis period. Your willingness to check parts of the text and brainstorm with me, along with the out-of-the-box questions that made me look at things from different perspectives, contributed more to this thesis than you probably realise.

Contents

List of Figures

List of Tables

List of Abbreviations

BERT	Bidirectional Encoder Representations from Transformers
CDP	Carbon Disclosure Project
CHF	Swiss Franc
CSRD	Corporate Sustainability Reporting Directive
CV	Coefficient of Variation
CZK	Czech Koruna
DKK	Danish Krone
EC	European Commission
ESG	Environmental, Social, and Governance
ETS	Emissions Trading System
EUR	Euro
GRAT	Greenwashing Risk Assessment Tool
GW	Gigawatt
IEA	International Energy Agency
IQR	Interquartile Range
LLM	Large Language Model
MAD	Mean Absolute Deviation
MW	Megawatt
MWU	Mann-Whitney U (test)
NGO	Non-Governmental Organization
NLP	Natural Language Processing
NOK	Norwegian Krone
PCG	Performance-Communication Gap
PLN	Polish Złoty
PV	Photovoltaic
RE	Refinitiv Eikon (dataset/stream)
$\bar{\boldsymbol{R}}_{\boldsymbol{S}}$	Ranking Shift
RQ	(sub-)Research Question
SBTi	Science Based Targets initiative
TF-IDF	Term Frequency-Inverse Document Frequency
TWh	Terawatt hour
USD	United States Dollar

1 Introduction

This introductory chapter outlines the foundation for developing a greenwashing risk assessment tool for electric utilities within Europe. Section 1.1 examines the climate finance gap and regulatory pressures driving this research need, while Section 1.2 reviews existing greenwashing detection approaches and their limitations. The chapter then identifies research gaps, presents the research questions, and explains how this work contributes to greenwashing detection methods.

1.1 Problem Statement

Energy Transition Context and Climate Finance Gap
If all net-zero promises and contributions were fulfilled, global warming could be limited to just under 2°C, which would reduce risk but still carry the danger of triggering irreversible climate tipping points [7]. But if present patterns continue, global warming is expected to approach +2°C by the late 2020s; recently it already exceeded 1.5°C above pre-industrial levels [17]. *"The energy sector is the source of around three-quarters of greenhouse gas emissions and holds the key to averting the worst effects of climate change, perhaps the greatest challenge humankind has faced"* [113].

Investment requirements reveal just how massive this challenge really is. Clean energy spending needs to reach $4.5 trillion per year by 2030 to maintain a realistic chance of limiting global warming, according to the International Energy Agency (IEA) [245]. The unfortunate reality is that current progress remains far off track: in 2024, clean energy investment was $2.2 trillion, only 37% of what is needed for net zero by 2050 [23]. This $3.3 trillion annual shortfall represents the gap between climate promises and climate action, making reliable sustainability information absolutely critical for smart investment decisions.

Electric utility companies find themselves right in the middle of it all. With energy sectors, especially focused on electricity generation, producing almost 25% of greenhouse gas emissions in Europe (shown in *Figure 1.1*), making them second only to transport in terms of sectoral contribution. Utilities face unprecedented pressure to transform their business models while maintaining energy security. This positions them as both major contributors to the climate problem and essential players in finding and creating a solution.

The electric utility sector is expected to be a pillar of net climate neutrality by 2050 [72]. However, due to digitisation and growing electrification of transport and heating, electricity demand continues to rise across Europe [134]. This growing electricity demand puts utilities in a difficult position during the energy transition. They must expand renewable capacity, phase out fossil fuels, and meet rising demand all at once [6].

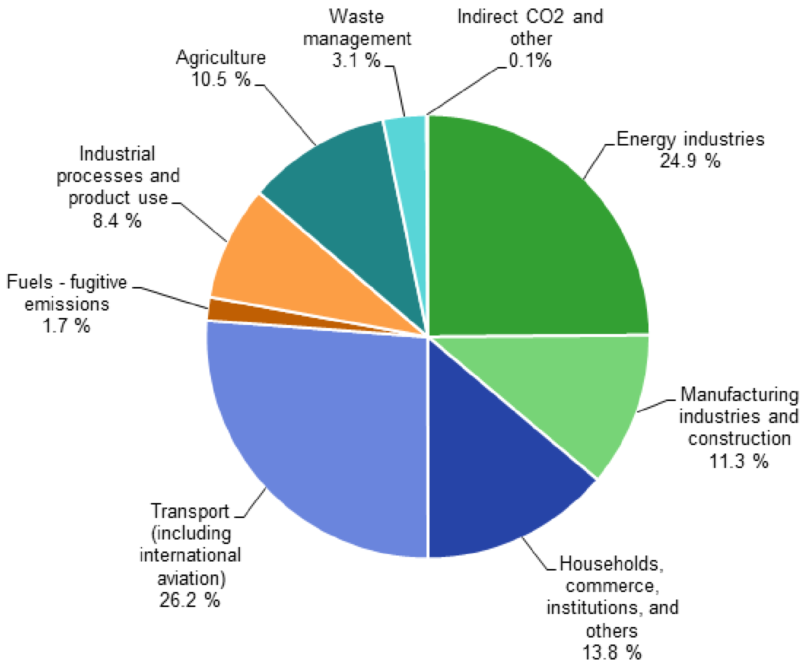

Figure 1.1: Greenhouse emissions in Europe per sector [56]

Specific Challenges in the European Utility Sector

Electric utility companies show a tendency toward greenwashing during periods of regulatory pressure, where companies overstate environmental commitments while continuing carbon-intensive operations [125]. The electric utility sector faces unique challenges that make detection of similar greenwashing cases particularly critical. Unlike other industries, electric utilities operate in highly regulated environments where consumer trust is fundamental to social acceptance of necessary infrastructure investments [191]. The EU single market creates challenges for utilities because cross-border electricity trading implies that when a company makes sustainability claims in one member state, these claims affect how credible the market looks across all 27 countries [69]. EU consumer protection laws go further than those in other regions, with *Directive 2005/29/EC* covering environmental claims and how they can mislead consumers [68]. The EU often leads the way in green regulation, and this "Brussels Effect" means that European utility greenwashing cases have influence far beyond Europe's borders [27].

Transparency in this sector has grown significantly as consumers often choose their electricity providers depending on their sustainability promises [213]. The sector is under pressure to integrate sustainable principles into its operations [222]. This creates conditions ripe for potential greenwashing as companies might oversell environmental progress to stay competitive and keep regulators satisfied.

Corporate responsibility in environmental sustainability has become more important given the pace of climate change and the need for systematic transformation [184]. However, a recurring challenge is the deceptive greenwashing tactic used by businesses to improve their public image by either exaggerating or falsifying their environmental initiatives [88]. When greenwashing is exposed, companies risk damaged reputations, legal action, and a loss of trust from investors and consumers [196]. For utilities, where business success depends heavily on public trust, these risks become particularly dangerous. European cases show that these risks are real. Under the EU's Green Claims Directive, penalties can hit up to 4% of a company's annual turnover, with member states able to set higher fines for repeat offenders [74]. Finland's Consumer Ombudsman found Vattenfall's claim *"fossil-free life within one generation"* misleading because it could not be proven true at the time of marketing [10]. In Germany, a court made a landmark ruling that major emitters like RWE can be held civilly liable for climate harms, showing European courts are taking these cases seriously [63].

Regulatory Urgency and Stakeholder Pressures
European regulation has made reliable greenwashing detection tools not just helpful, but necessary. The Corporate Sustainability Reporting Directive (CSRD) started in 2024, with the first wave of companies already publishing sustainability reports in 2025 covering their 2024 data [65]. Large utility companies that were not reporting before, start in 2026 with their 2025 numbers, and sector-specific standards arrive by June 2026 [28]. This timeline can be seen in *Figure 1.2* below.

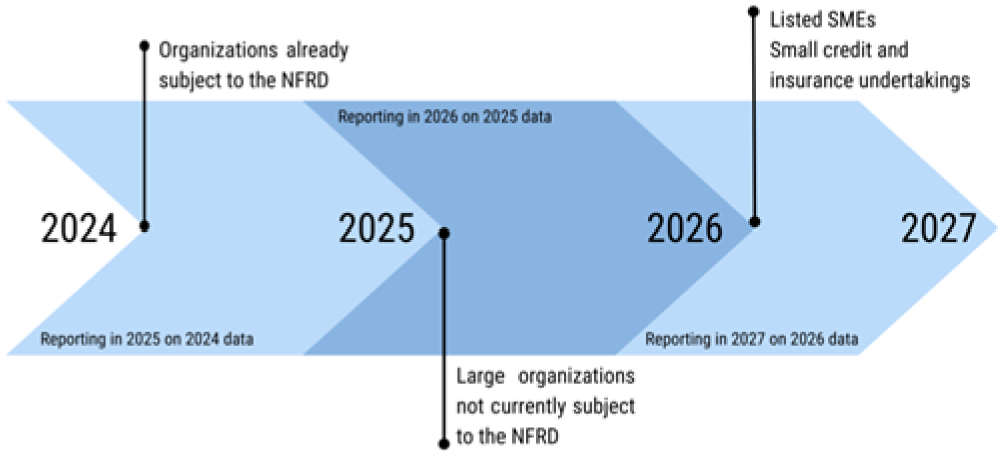

Figure 1.2: CSRD timeline [145]

Beyond CSRD, utilities must deal with the EU Taxonomy Regulation, which requires them to prove they make a substantial contribution to climate objectives and demonstrate *'do no significant harm'* to other environmental goals [194]. This creates binding technical screening criteria that utilities must meet to qualify as *'green'* investments. Enforcement varies across member states, with the EU requiring businesses to *"support claims with scientifically accepted evidence"* while banning unsubstantiated aspirational claims [67]. The new Directive 2024/825, which started in March 2024, lets member states set up enforcement measures including legal actions for companies that fail to follow environmental claim standards [66].

The regulatory pressure comes at the same time that stakeholders are demanding real accountability. Investors increasingly depend on solid environmental data for their decisions. When companies lie about their green efforts, investor trust crumbles, stock prices decrease and future funding becomes harder to secure [241]. On the consumer side, misleading environmental claims create frustration and scepticism that spreads far beyond the guilty company, making life harder for genuinely sustainable businesses [231].

Given the mentioned challenges and regulatory scale, traditional greenwashing detection methods face significant limitations. Manual analysis of sustainability reports, takes a lot of time and produces inconsistent results when applied across multiple companies [87]. The scale of European sustainability reporting makes this approach impractical for systematic oversight. Under the CSRD, approximately 50,000 EU businesses must provide sustainability disclosures, which is more than four times the roughly 11,000 companies covered under the previous directive [179]. These companies produce thousands of pages of sustainability reports annually, creating a massive analysis challenge. This situation creates an urgent need for automated, standardised detection methods that can process large volumes of text efficiently [20].

1.2 Literature Review

Despite the urgent need for greenwashing detection in today's landscape, the concept greenwashing itself was introduced a long time ago. Jay Westerveld actually coined the term greenwashing back in 1986, targeting corporate policies that misled consumers about environmental activities [48]. Westerveld called greenwashing *"a method of increasing profits, rather than [...] genuinely engaging in environmentally friendly activities"* [219]. The concept did not gain much traction in academic circles until Greer and Bruno [98] wrote about it in their environmental marketing book [209]. From there, the definition kept expanding to cover all sorts of deceptive corporate practices around environmental claims.

Greenwashing Definitions and Theoretical Development
An important problem is that there is still no definition on which everyone agrees. This has left the research field scattered and uncoordinated, regulations unevenly applied, and enforcement inefficient [219]. When researchers cannot agree on what greenwashing actually is, it becomes difficult to detect it consistently across different industries and countries. Most definitions get stuck on whether companies meant to deceive, which misses cases where companies are just defending themselves against unfair criticism [209]. Without a universal definition or standard detection methodology, allegations often remain inconclusive [173].

Freitas Netto et al. [84] tried to tackle this problem by creating four categories: firm-level claim, firm-level executional, product-level claim, and product-level executional. The firm-level claim type is subtle and easy to miss because companies make big environmental promises without backing them up with real action [141]. Spaniol et al. [219] took this further and identified six key elements: environmental focus, firm origin, consumer marketing, unsubstantiated claims, deceptive intent, and competitive advantage.

The definition that most academics have settled on comes from Delmas and Burbano [48]. They define greenwashing as *"the intersection of two firm behaviours: poor environmental performance and positive communication about environmental performance"*. This works well because it focuses on the mismatch between what companies say and what they actually do, plus researchers can actually test it [150]. Their framework shows two main ways in which greenwashing occurs: decoupling (creating symbolic policies without real change) and selective disclosure (showing off the good while hiding the bad). These ideas help explain how companies keep their green reputation while performing poorly on environmental issues.

Performance-Communication Research Approaches
Greenwashing research has split in two, with little exchange between the different sides. One side studies environmental performance using hard numbers like carbon emissions, resource use, or regulatory violations [43]. The other side analyses what companies actually say/claim through content analysis, sentiment studies, or surveys about stakeholder perceptions [159]. Bringing these approaches together rarely happens. Most studies pick either performance or communication, not both. When performance data shows up in communication studies, it is usually just background information. When performance studies mention communication strategies, they do not actually analyse the messages. This split makes sense from a methods standpoint since it is hard to combine quantitative environmental data with qualitative text analysis.

Some recent work has started bridging this gap. Lagasio [130] created an Environment, Social, and Governance (ESG)-Washing Severity Index using keyword counts and sentiment from CSR reports. She found differences across industries and regions, which is useful for measuring communication tone, but her approach ignored actual companies performance data. Bindi [18] compared climate messages on corporate websites with how companies actually lobby at the EU level. This revealed political contradictions but did not check whether companies actually hit their stated climate targets. CDP stands out as a well-structured and widely used source for emissions, climate targets, and verification data [90]. Other sources, like third-party assurance and audit-based databases, also contain useful emissions and target data. Even so, most greenwashing research either overlooks these resources or only uses parts of it [81].

NLP Applications in Environmental Communication

Natural Language Processing (NLP) has become a game-changer for analysing corporate environmental communication at massive scale. Early work mostly used sentiment dictionaries or keyword counting to evaluate CSR report tone [120, 77]. These methods revealed that environmental communication tends to be overly positive, but they missed deeper language patterns. Recent advances have moved way beyond basic sentiment analysis. Wang, Gao, and Sun [238] built a BERT-based (Bidirectional Encoder Representations from Transformers) model that scores individual sentences for greenwashing based on narrative structure, exaggerated sentiment, and how specific the language is. Their work demonstrates how deep learning can catch subtle communication patterns that simple polarity analysis misses. Shimamura, Tanaka, and Managi [211] used large language models (LLM's) to evaluate how clear sustainability reports actually are. They found that companies accused of greenwashing wrote reports that were less logical and harder to read than their peers.

Topic modelling and semantic embeddings allow for more sophisticated analysis of communication themes. Boelders [24] used these techniques to connect sustainability claims with green technology patent areas in European energy companies. This made it possible to cluster communication and innovation domains without supervision. Saxena [202] pushed this approach further by combining claim-specific NLP indicators with actual emissions to check the credibility of climate-focused investment portfolios.

But major methodological problems remain. Calamai et al. [31] point out that there is still a lack of unified benchmarks or datasets for greenwashing detection. Climate-specific language models like ClimateBERT are starting to appear, but a lack of labelled data, clashing definitions of greenwashing, and weak regulatory coordination continue to slow down real progress. Most NLP approaches focus on just one aspect of communication quality instead of the multi-dimensional analysis needed to capture how complex corporate environmental messaging really is.

Greenwashing in European Electric Utilities

Electric utilities in Europe occupy a special spot in greenwashing research. They are central to energy transition, heavily regulated, and large emitters [131]. Unlike consumer-facing industries where greenwashing mainly affects buying decisions, utility sector claims directly shape energy transition success and climate policy effectiveness [232]. These companies operate under extensive regulatory frameworks including the EU Taxonomy (CSRD), and Emissions Trading System (ETS). This creates both restrictions and incentives for environmental communication.

Despite this strategic importance, most empirical greenwashing studies focus on high-visibility consumer sectors. Industries such as fashion, technology, and retail often get the most attention, since their green branding is directly visible to consumers [84, 77]. The utility sector gets limited targeted analysis. However, a few exceptions exist. As mentioned before, Boelders [24] analysed how CSR and patents line up in European energy companies using NLP techniques. His study showed that energy companies often promote certain green technologies more in their messaging than their actual innovation efforts support. But this analysis used patent data as a stand-in for environmental action rather than looking at operational performance outcomes. Bindi [18] included large utilities in the cross-sector study of communication-lobbying mismatches but did not separate utility-specific behaviour patterns or account for sector-specific regulatory constraints.

Cross-sector studies suggest that regulatory design affects disclosure practices. Luu et al. [139] found that mandatory emissions reporting can reduce greenwashing behaviour. This indicates that regulatory frameworks do limit deceptive communication. However, a systematic analysis of how European utilities communicate is still largely missing. Because of its regulated structure, long-term investments, and public service role, [46] the sector faces distinct stakeholder demands that could lead to communication patterns different from those in other (previously mentioned) competitive industries.

1.3 Research Gap

Three major gaps reveal fundamental limitations in greenwashing theory that prevent the development of systematic detection, especially for European electric utilities.

Performance-Communication Integration Gap
Most greenwashing literature treats performance and communication as completely separate theoretical topics. Studies focused on performance analyse emissions, resource use, or regulatory compliance. Communication studies focus on message content, tone, or stakeholder perception without linking it to actual environmental outcomes. This represents a core limitation in greenwashing theory: existing frameworks cannot explain how performance-communication relationships actually operate. While studies like Saxena [202] attempt integration, they lack systematic methodologies for combining quantitative performance data with detailed communication analysis. No research to date has developed a practical tool for assessing greenwashing across multiple dimensions beyond the basic performance-communication gap.

Multi-Dimensional NLP Analysis Gap
Most existing research has focused on single-dimensional analysis, such as sentiment scoring or keyword frequency. This limited focus fails to capture the layered communication strategies that define modern greenwashing. As a result, it faces a key theoretical limitation. It cannot capture how misleading signals are coordinated across different dimensions simultaneously. Current methods assess communication aspects in isolation, without integrating them into a broader quality framework. Climate-specific tools such as ClimateBERT also face limitations due to uneven definitions and training datasets. Most importantly, the ways in which communication dimensions combine to influence greenwashing risk remains unexplored.

Sector-Specific Analysis Gap
Although some studies have analysed aspects of greenwashing in the European electric utilities sector, many areas remain overlooked despite the sector's central role in the energy transition. Most research focuses on consumer-oriented industries, leaving theoretical gaps in how greenwashing appears in regulated, capital-intensive sectors. Electric utilities operate under specific stakeholder demands and communication limits, which likely result in different forms of greenwashing. Even studies that include these utilities, such as Boelders [24], treat them as part of the broad energy sector rather than as an independent industry group. This gap points to a fundamental limitation in current greenwashing frameworks. The lack of systematic analysis on how regulatory contexts influence greenwashing is problematic given the role utility companies play in climate action.

These gaps show that greenwashing research lacks integrated systematic tools that can detect patterns across multiple analytical dimensions. Addressing these limitations creates important opportunities for academic advancement. This research makes an important contribution to greenwashing theory by introducing the first systematic integration of performance measurement and multi-dimensional communication analysis.

1.4 Research Questions and Structure

The literature gaps found point to a clear research opportunity: developing better greenwashing detection methods tailored for electric utilities companies within Europe. The sector provides a strong foundation for developing new methods while directly addressing the knowledge gaps mentioned above.

1.4.1 Research Questions

The **main research question** asks: *What greenwashing risk assessment tool can be developed for European electric utility companies based on environmental performance metrics and multi-dimensional communication analysis?*

The core research focuses on this question, it responds to the performance-communication integration gap by explicitly requiring both dimensions. It also targets the sector-specific analysis gap by focusing on utilities, and the multi-dimensional NLP gap by requiring analysis across multiple communication quality dimensions. The main question is supported by three sub-questions, each targeting a different aspect of detecting greenwashing. Together they form an integrated approach that tackles the methodological problems found in existing literature.

Three **sub-questions**:

RQ1: *What environmental performance variations exist among European electric utility companies based on systematic analysis of their available sustainability data?*

The first sub-question establishes the performance baseline needed for greenwashing detection. It tackles the current research tendency to either ignore performance data entirely or use it only for basic sorting. By examining performance variations systematically, this question creates the foundation for spotting performance-communication mismatches.

RQ2: *What multi-dimensional communication patterns characterise environmental disclosure in European electric utility companies?*

The second sub-question attacks the multi-dimensional NLP analysis gap by examining communication across several quality dimensions rather than focusing on single metrics like sentiment. It also contributes sector-specific knowledge by identifying communication patterns unique to utilities rather than applying frameworks or tools built for consumer-facing industries.

RQ3: *What greenwashing risk patterns are revealed when environmental performance and communication characteristics are analysed together?*

The third sub-question addresses the difficulty of integration by systematically combining performance and communication analysis. Instead of just identifying poor performance or misleading communication separately, it looks for the specific combinations that signal greenwashing risk.

1.4.2 Research Structure

The remainder of this thesis is organised into six chapters following the systematic approach outlined in the research flow diagram (*Figure 1.3*). The conceptual framework in Chapter 2 lays out the theoretical foundation that guides the analysis, as shown in Phase 1 of the research. It draws on performance-communication theory and regulatory compliance frameworks to build the theoretical foundation for integrated greenwashing detection.

Chapter 3 Methodology explains the methodology in detail, implementing the data collection and analysis approaches from Phases 2 and 3. This covers data collection procedures for both performance metrics and corporate communication content. It also explains the multi-dimensional NLP analysis approach and other methods used to spot greenwashing risk patterns. Results presented in Chapter 4 show the empirical results. It starts with performance variation analysis across the different electric utility companies, then moves to communication pattern identification, and finishes with integrated greenwashing risk assessment findings as outlined in the tool's progression from individual analyses to integrated risk assessment. Chapter 5 focuses on practical application of the GRAT. This is a standalone chapter meant for providing guidance on how users can apply the tool to assess greenwashing risk within the electric utility sector and adapt it for other industries. This includes step-by-step implementation processes, data requirements, sector adaptation methods, and validation approaches that users can apply for their specific contexts. Chapter 6 discusses the findings through two perspectives: first examining the GRAT as a developed tool, including its limitations, user requirements, and adaptability features, then reflecting on the patterns observed among the 14 companies analysed and their broader implications within the acknowledged sample constraints. Chapter 7 concludes the research by directly answering each research question, reflecting on the tool's key characteristics, and outlining clear directions for future research and practical implementation of greenwashing detection methods.

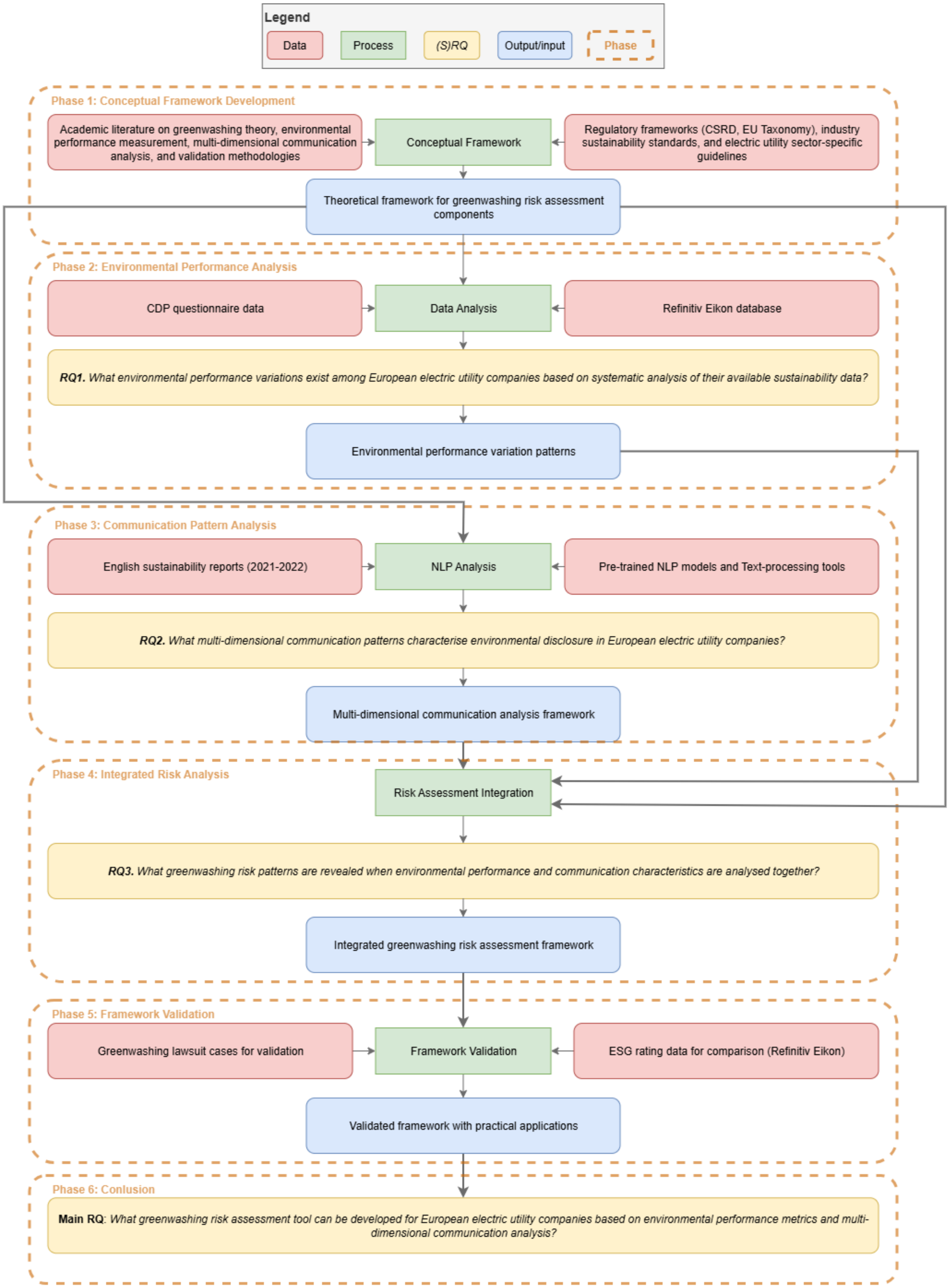

Figure 1.3: Research flow diagram

1.5 Research Positioning and Contribution

This research tackles the need for systematic greenwashing detection specifically for electric utilities companies within Europe. The utilities sector accounts for the highest relative occurrence (15.91%) of greenwashing cases [123]. This sector is suited for developing a new detection method because utilities are within a regulated sector that faces unique transparency requirements that setting them apart from other industries. Second, their central role in energy infrastructure means their sustainability claims directly impact climate transition success. Third, their standardised reporting through frameworks like the Carbon Disclosure Project (CDP) provides comparable data for systematic analysis [33].

Greenwashing reduces the effectiveness of climate action and slows down assessment of environmental progress [227]. More than half of the evaluated company websites displayed practices connected to greenwashing [72]. This misinformation not only affects consumer decision-making but also policy-making efforts, as companies' misleading practices can hide their environmental impact [19]. In this context, detecting greenwashing becomes important for protecting financial systems and promoting transparency in corporate sustainability. While the importance of detecting greenwashing is clear, the process itself remains complex. Definitions of greenwashing differ greatly and often includes corporate practices, marketing, and sustainability reporting [84]. This lack of standardisation exposes a gap in the current approaches and makes greenwashing challenging to detect. Current methods mostly rely on manual, qualitative assessments, which are time-consuming and often fail to capture the subtle interactions of greenwashing in context with a lot of data [88].

This thesis addresses the mentioned challenges by concentrating on electric utility companies within Europe, as the sector is a major contributor of emissions but also one with great potential for decarbonisation. It develops a comprehensive Greenwashing Risk Assessment Tool (GRAT, from here on) that integrates quantitative environmental performance metrics (derived from corporate emissions data) with multi-dimensional NLP-based communication. This approach responds to the growing need for new methods to systematically tackle greenwashing as corporate sustainability claims increase [73]. The goal is to give consumers and policymakers useful knowledge so they can identify real sustainability initiatives among false claims. Furthermore, it addresses the larger social implications of tackling greenwashing, which is required to build confidence in green markets and satisfy world climate targets [187]. Reaching this target calls for active, open participation of every involved actor, especially companies.

2 Conceptual Framework

This chapter develops the theoretical foundation required to create a systematic greenwashing detection tool. *Figure 4* illustrates how four foundational theories (Section 2.1) inform the development of domain-specific theoretical frameworks for performance measurement (2.2), communication analysis (2.3), and computational methods (2.4), which then integrate through validation theory (2.5) to enable practical implementation. Each theoretical component addresses specific aspects of greenwashing detection while contributing to the overall GRAT architecture.

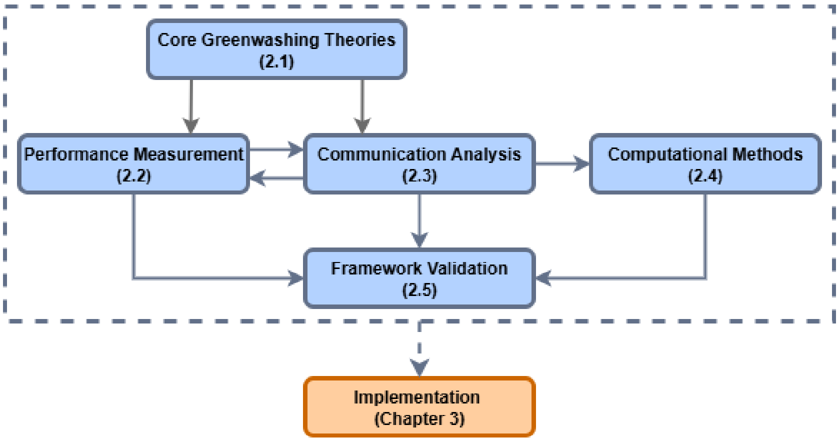

Figure 2.1: Conceptual Framework Development

2.1 Theoretical Foundations of Greenwashing Detection

This research adopts the definition by Delmas and Burbano [48], which, as discussed in the literature review, is the one most commonly used definition of greenwashing in academic research: *"the intersection of two firm behaviours: poor environmental performance and positive communication about environmental performance"* [48]. This definition captures greenwashing not simply as a matter of misleading claims on its own, but as a measurable disconnection between real environmental behaviour (e.g. carbon emissions) and how that behaviour is presented to the outside world. The gap between these two forms the foundation of the approach taken in this research. Four theories explain different aspects of this gap:

2.1.1 Decoupling Theory

According to decoupling theory, organisations may present themselves as environmentally committed while avoiding actual change within their operations. Meyer and Rowan [155] first described this process as a way for companies to shield their formal claims from technical uncertainty by adopting widely *"accepted symbols and structures"* that reflect external expectations. These elements help organisations to appear credible, even when their actual practices do not fully match (are decoupled from) what they formally claim [155]. In sustainability contexts, companies may adopt reporting standards or set long-term climate goals. At the same time, their daily decisions often follow older, less sustainable practices. Bromley and Powell [29] build upon this concept, showing that decoupling can become systemic, especially when stakeholders focus more on appearances than on actual results.

This research uses decoupling theory as the key rationale for connecting environmental communication to real-world performance data.

2.1.2 Information Asymmetry

This separation persists because of information gaps. The concept of information asymmetry, originally introduced by Akerlof [4], shows how limited access to information can affect the quality of decision-making. In the context of environmental reporting, this means companies have more knowledge about their operations than outside stakeholders such as consumers and regulators. As a result, companies can focus more on positive instead of less favourable or negative details. Kashyap, Verma, and Pal [122] found that companies with weaker environmental performance are often less transparent in their disclosures. This opens the door for greenwashing to grow and go undetected. This research uses that theory to justify using both self-reported and independently sourced emissions data. This can uncover reporting gaps or subtle signs of greenwashing that would be hard to spot in isolation.

2.1.3 Stakeholder Theory

Companies exploit these gaps differently depending on their audience. The stakeholder theory [82] offers a useful perspective. Companies face different expectations from customers, regulators, and investors. These groups require different information and they do not all interpret environmental performance in the same way [133]. Companies may respond by adjusting how they communicate, focusing on what sounds convincing to each audience rather than what fully reflects the facts. Stakeholder theory helps explain why companies present sustainability messages in layered and selective ways. It also supports analysing both the content and the quality of communication.

2.1.4 Signalling Theory

These strategic choices reveal themselves through language patterns. Signalling theory helps to explain the logic behind some of the language choices found in sustainability reporting. Spence [220] originally developed the concept to describe how individuals and organisations use signals to communicate qualities that cannot be directly observed. In environmental reporting, these signals include carbon neutrality claims, science-based targets, or polished report formats [150]. Companies can use these tools to present a green image without making real progress [228]. That possibility is especially relevant for this research, which considers not just the volume of sustainability communication, but its substance and clarity. Signalling theory justifies looking beyond sentiment and frequency and focuses on whether claims are supported by concrete evidence.

Together, these theories justify examining both performance data and communication quality to detect greenwashing.

2.2 Performance Measurement

Building on the theoretical foundations that established greenwashing as a gap between performance and communication, this section develops the framework for measuring actual environmental performance. The previous chapter showed why companies might decouple their claims from reality. The next challenge is finding reliable ways to measure what they actually do.

2.2.1 Measurement Challenges

Evaluating a company's overall environmental performance is more complicated than it seems. Emission data should sit at the heart of environmental assessment [54]. But environmental disclosures often lack credibility, since managers control both the content and assurance process, resulting in limited transparency [176]. There is a paradox here: companies are the main source of environmental data, but their reports cannot be fully trusted.

Triangulating multiple data sources offers a possible solution. CDP has become the main source for climate disclosures globally [118]. But there are checks and balances needed. Third-party validation makes data more objective and reliable for academic research [132, 37]. Refinitiv Eikon (RE) stands out due to its reliability and transparency, which make it suited for academic ESG studies [214]. When discrepancies appear, proportional penalty models offer a fair way to handle the differences [75, 188].

2.2.2 Performance Components

The research literature points to five important dimensions for measuring environmental performance.

Emission intensity comes first. Major indices like MSCI and CDP prefer intensity metrics (emissions relative to revenue or operation) over raw emissions because bigger companies naturally produce more emissions [205, 78]. There is a need for standardised metrics for fair comparison, with tracking and industry benchmarking [151, 54].

Target assessment tells only part of the story. Companies set targets that matter across multiple dimensions: type, timeframe, and ambition level [55, 118]. Kemp, Pontikes, and Hawn [124] make an important point: ambitious goals need transparency and accountability to be credible. Annual tracking helps connect goals with accountability and established goal-setting theory [233, 185]. This creates two separate measures: how well companies achieve existing targets (backward-looking) and how ambitious their future targets are (forward-looking).

Transparency addresses information consistency across different environmental data sources. Transparency drives actual environmental improvement because stakeholders with better information can hold companies accountable for environmental performance [223]. When used as a governance tool, transparency can push companies to cut back on environmentally harmful practices [22].

Renewable energy deserves attention for electric utilities. Adopting renewables improves environmental performance and plays a central role in solving the energy crisis [101, 127]. In this sector, shifting to renewable energy represents the most direct route to cutting emissions. Renewable performance should cover multiple areas, supporting extra credit for companies making real transitions [183].

Suspicious target changes raise a concern that is related. Vague or shifting targets serve as greenwashing strategies [231]. Companies frequently modify or soften ESG commitments [159] and remove targets without explanation [118]. These missing data patterns link with poor overall reporting quality [38].

2.2.3 Methodological Framework Approach

The literature gives us clear direction for putting each component into practice.

Emission intensity brings its own challenges. Outliers can throw off an analysis, as extreme values lead to systematic bias in the data [129]. Trimming these extreme values resolves this [181]. The scoring needs to reflect reality: companies hitting really high performance should not get massively higher scores than good performers. Square root functions work well for this [14].

Target achievement needs careful handling. Progress tracking is a central part of good performance evaluation. Metric-based methods give us a systematic way to measure performance against stated goals [200].

Target ambition assessment gets more interesting when there are real benchmarks. Verde et al. [234] use the Science Based Targets initiative (SBTi) standards to assess European companies, SBTi shows exactly what annual reduction rates companies need for 1.5°C alignment [206, 207]. Some companies game the system by inflating their baselines [249], which exaggerates their progress. Using recent-year performance as a reference provides a more accurate picture. Short- and long-term targets should be treated separately, and normalisation in some form helps with creating fair comparisons [21]. When translating this into scores, linear scaling helps maintain clarity and transparency [13].

Transparency can be measured by comparing self-reported and third-party verified emissions data. As mentioned earlier, CDP could be a key source of self-reported data; RE could serve as the third-party verified counterpart. Large gaps between these may indicate selective disclosure [132, 174]. Minor differences may indicate different scope definitions or calculation methods [195]. Proportional scoring should be applied and based on the magnitude of the discrepancies [75, 188].

Renewable energy metrics follow principles found in performance management approaches. Dual frameworks are most effective, mixing structural factors like renewable capacity with procedural ones such as investment commitments [80]. Absolute measures do not work well here. Comparative benchmarking provides a more practical approach, especially when data availability is limited [53]. Within consistent groups such as industry sectors, relative grading, similar to approaches used for emission performance works well [8]. The scoring should encourage improvement [237], with effective metrics being challenging but achievable [226].

Suspicious target changes require penalty scoring based on growing signs of reduced credibility [221, 36]. Companies that rely on deceptive environmental claims, including target manipulation, face increased regulatory penalties as transparency standards evolve [62].

2.2.4 Integration Considerations

Temporal aspects are very important when considering performance data, tracking changes over time is always more revealing than looking at isolated moments [78]. It reveals trends and patterns that single-year data cannot capture. Both spatial and temporal tracking improve reliability [239]. This supports examining multiple years of data for performance trends.

2.2.5 Scoring and Weighting

Theoretical foundations support integrating performance components through structured approaches. Relative scoring works better than absolute measures for fairness [198]. Companies may manipulate targets to appear more successful when performance and goal-setting become "tightly coupled", a behaviour known as sandbagging [126]. Breaking these dimensions apart helps prevent this problem. Isaacs et al. [115] found something unexpected: ambitious goals actually reduce sandbagging. This supports keeping target ambition separate from achievement.

Literature reveals clear component hierarchies. Emissions data sits at the centre of environmental assessment, showing emission intensity plays the primary role [54]. Target achievement connects to accountability theory [103]. Target ambition reflects commitment quality. Transparency supports other dimensions but should not overshadow direct performance measures [51].

Weight assignment creates substantial challenges in composite environmental indicators [14]. Weight effects are often misunderstood because correlation patterns among variables cause actual indicator importance to differ from intended weights. This weight uncertainty needs a systematic solution. Instead of picking arbitrary weights or using random sampling, constrained ensemble scoring generates every valid weight combination within reasonable bounds [106]. This covers the entire feasible weight space while respecting theoretical priorities through designed constraints. Each component gets bounded within sensible ranges, and hierarchy is connected to the theoretical relationships [201].

Systematic enumeration beats making single arbitrary decisions about uncertain parameters [92]. Constraint-based methods handle multiple theoretical priorities, avoiding claims of precision that are not accurate [106]. Final scores combine all valid options, showing methodological uncertainty instead of hiding it.

Bonus and penalty mechanisms need proportional scaling principles. These follow best practices that recommend penalty-reward structures "correspond to weakness and effectiveness indices" with proportional scaling [36]. Preventing both inadequate incentives and unstable score fluctuations while maintaining interpretability ranges. Multiple sources support structured weighting approaches when theoretical priorities exist [9]. Unequal weights work when priorities have explicit justification and standardised scoring ranges support clarity [25].

2.3 Corporate Environmental Communication

The previous section established how to measure how companies perform environmentally. This section will find theories that describe how companies communicate about their environmental performance. This is more than only identifying untrue claims. It is about understanding the subtle techniques companies use to influence perception through language use, timing, and repetition.

2.3.1 Theoretical Foundations for Communication Analysis

Corporate environmental communication goes beyond simple PR work. Sustainability reports have become an important way in which companies formally share environmental information with investors, regulators, and consumers [128]. Unlike marketing materials, these reports face legal requirements, particularly under frameworks like CSRD and EU Taxonomy, which demand standardised, verifiable disclosures [177]. This regulatory shift turns sustainability reports into legally important documents, making them suited for academic and policy analysis [175].

Advanced corporate communication analysis needs more than only simple keyword counts or sentiment measures. Keyword frequency and sentiment tools often miss message framing, temporal signals, and consistency over time [121]. Researchers stress that message framing matters because how companies present issues differently affects stakeholder interpretation [52]. The timing and frequency of sustainability commitments shape how credible they seem and how engaged audiences are [154]. This evidence shows that examining framing, timing, and consistency together captures the complexity of corporate environmental communication better than basic content counts.

2.3.2 Performance-Communication Gap Foundation

Greenwashing happens when symbolic communication does not match actions [236, 84]. Companies engage in selective disclosure, highlighting only the good news while hiding the less favourable details [140]. When companies dramatically increase their environmental communication while their performance remains poor, this reflects greenwashing in its most problematic form [5, 141].

2.3.3 Multi-Dimensional Communication Analysis

As established in the literature review and Chapter 2.1, the gap between actual performance and communication forms the core of greenwashing. But there is a problem: while the Delmas & Burbano's framework tells us what greenwashing is conceptually, it does not explain how to measure “positive environmental communication” in practice. This creates an operational challenge.

The literature reveals the need for multiple components. First, there is the mistake of focusing on just one aspect. Nemes et al. [173] identified a critical gap: "we have not found any attempt in the academic literature to give weights to the different indicators in greenwashing frameworks." This suggests existing approaches oversimplify greenwashing as a single phenomenon when it's actually multi-faceted. Second, positive environmental communication itself has multiple dimensions. Regulatory frameworks require companies to support claims with scientifically accepted evidence while restricting unsupported aspirational claims [76, 71]. The EU Green Claims Directive requires "robust, science-based evidence" while prohibiting vague environmental claims [67]. These requirements show that measuring communication quality involves multiple distinct elements. Third, certain communication patterns serve as independent greenwashing indicators regardless of performance levels [76, 231, 117]. Finally, composite indicator best practices support avoiding single-indicator dominance to prevent misleading scores [93].

2.3.4 Communication Strategies

Research identifies five distinct communication strategies that companies use to manage environmental perception, as shown in Table 2.1 below.

Table 2.1: Communication strategies found in literature

Dimension	Description
Sentiment manipulation	Emphasising positive emotional language around environmental topics.
Evidence manipulation	Strategic use of substantiated "hard" disclosure versus unsubstantiated "soft" disclosure.
Language vagueness	Using vague or meaningless terms and avoiding concrete commitments; however, appropriate hedging may also reflect authenticity.
Temporal manipulation	Making distant promises about environmental goals without including near-term action plans.
Template replication	Repeating similar language or structure across years, allowing symbolic compliance without substantive communication changes.

These five strategies require distinct measurement approaches to capture their sophisticated implementation.

Performance-Communication Gap
This forms the foundation of greenwashing theory [236, 84]. Research shows this gap approach works better for detecting greenwashing than alternative methods, with empirical evidence demonstrating superior performance [15, 227].

Environmental communication theory identifies two core elements: sentiment framing and green communication volume. When companies discuss environmental issues positively, it directly affects stock prices and correlates with ESG ratings [16, 250]. Renewable energy and climate topics carry distinctly positive sentiment that shapes public perception [135]. Research shows that "receptivity to green communication was the most important factor affecting green attitudes" [248]. It focuses on "sustainability messages" and environmental content coverage, examining topics like climate, energy, and sustainability [89].

Communication volume is grounded in selective disclosure theory, where companies reveal positive environmental information while hiding negative aspects [140]. This mechanism requires sufficient environmental volume density and vocabulary diversity to work because companies cannot selectively disclose what they avoid discussing entirely. The approach captures the core greenwashing definition by combining green term density and sentiment with actual performance data. Adomßent and Godemann [2] showed sentiment matters more than frequency for public impact.

Multiple empirical studies support this hierarchy across different contexts. Analysis of 42,000 news headlines shows that sentiment scores shift considerably as time progresses independent of message frequency [30]. Experimental data from environmental email campaigns reveal that messages with negative sentiment generate much greater engagement, regardless of message frequency [105].

However, term frequency and vocabulary diversity both play a role in selective disclosure effectiveness. Studies show that coverage and frequency of lexical items tend to predict information processing better than vocabulary diversity, supporting frequency-based approaches [35]. Disclosure credibility also increases through frequency and consistency rather than vocabulary diversity [247]. Vocabulary diversity remains important, as it can shape how stakeholders interpret environmental disclosures by signaling the effort and engagement behind communication [180, 119].

In cases where performance-communication gaps (PCG from here on) show low performance together with green communication, applying amplification factors is necessary. According to Mariani and Ciommi [144], amplification factors help reveal imbalances in composite scoring by scaling relative variability. Following best practices for the robustness of the composite index, analysts often apply moderate amplification factors (commonly around 1.3–1.5) to assess the sensitivity of results to scoring thresholds [93].

Substantiation Quality
This operationalises Clarkson et al. [43]'s distinction between "hard" disclosure (numbers, evidence, verifiable facts) and "soft" disclosure (aspirations, promises, unverifiable statements), finding that "hard disclosure is more credible than soft disclosure." Cho, Roberts, and Patten [39] show that "worse environmental performers exhibit significantly more 'optimism' and less 'certainty' in their language." This linguistic pattern correlates with but exists independently of performance gaps.

This theoretical basis is supported by recent regulatory frameworks. The EU Green Claims Directive defines clear hierarchical requirements for quantified, science-based evidence over aspirational claims [70]. It requires "robust, science-based, and verifiable methods" supported by third-party verification. Quantitative disclosures are associated with greater clarity of information and investor trust, showing a preference for objective data over qualitative statements [156]. The Federal Trade Commission requires competent and reliable scientific evidence for environmental claims, with thorough substantiation needed for unqualified claims [76].

Language Clarity
Studies repeatedly show that vague language is a form of greenwashing that creates scepticism and damages authenticity [231, 84, 252, 203]. The pattern appears across industries, with vague wording such as "eco-friendly" and "green" that mean nothing concrete [209, 26]. Vague claims reduce consumer trust regardless of actual performance verification [57, 117].

But legitimate companies use cautious language to avoid over-promising. Hedging expresses appropriate uncertainty, which is essential for credible scientific communication [235, 111]. Li and Chen [136] show hedging correlates with credibility expectations across cultures. When companies choose their words carefully, they build stakeholder trust [153, 186, 204].

Separating vague language from legal hedging is empirically validated. Using four-metric models to distinguish vague language from legal hedging, Capetz et al. [32] achieved 86.34% accuracy in detecting greenwashing. This shows that hedging can indicate sincere environmental claims.

Temporal Orientation
Montgomery, Lyon, and Barg [159] identified companies making impressive net-zero promises for 2050 while not making any (recent) progress. The United Nations 2024 warns these distant claims deceive

without short-term plans. Without near-term goals, long-term commitments lack credibility [102]. Real commitment requires long-term pledges paired with detailed short-term actions and milestone tracking Oxford Net Zero [182] and UN High-Level Expert Group on Net Zero Emissions Commitments [230]. When companies fail to meet environmental promises, this links directly to greenwashing and legitimacy loss [100, 225].

Research on temporal framing supports the relative importance between focussing on the future and mentioning specific timelines. When companies operate in languages that clearly separate future and present tenses, they tend to invest less in future-focused activities like sustainability [137]. How companies communicate about the future affects what stakeholders expect, particularly when language does not match present actions. Messages that create fear work better when focused on the short-term, while hopeful messages need specific timelines instead of vague future promises [110]. Environmental messages that reference the past rather than the future boost both credibility and people's willingness to act, showing the importance of how much timeline reference [11].

Reporting Consistency
Choi, Chung, and Young [41] discovered companies systematically repeating structure and content across years, suggesting symbolic rather than substantive reporting. [240] analysed over 10,000 sustainability reports and found widespread boilerplate language that correlated with ESG rating problems. [42] found half of US companies use generic, repetitive language in ESG filings. Audits check facts but ignore how communication quality develops over time [107]. This allows companies to meet disclosure requirements through unchanged reporting while failing to show genuine progress.

Recent NLP research shows the need for separate cross-year similarity from sentence-level repetition. Kang and Kim [121] found that sentence similarity methods using Sentence-BERT models work much better than keyword matching for detecting actual content changes. Their 10-year analysis uncovered "monotonous thematic patterns" in sustainability reports, suggesting companies use symbolic rather than actual reporting. When analysing similarity across years, companies often repeat the same themes rather than advancing their environmental commitments. Sentence-level analysis then backs this up by revealing template or boilerplate-based communication.

2.3.5 Weight Allocation

Within each communication dimension, components need weights that reflect their relative importance. The problem is straightforward: academic literature gives no guidance for weighting indicators in greenwashing frameworks Nemes et al. [173]. This creates tension between theoretical precision and empirical reality.

Some directional priorities do exist in the literature. Sentiment matters more than frequency in environmental communication [2, 248]. Hard disclosure works better than soft disclosure [43]. Vague language poses greater greenwashing risk than legitimate hedging [32]. These studies show direction but not magnitude, for weighting.

Using structured weight patterns based on evidence strength offers a balanced approach. When literature shows clear but moderate preference, 60–40 splits reflect the hierarchy. Sentiment gets priority over frequency, but frequency is still important for selective disclosure theory [140]. Stronger evidence supports 70–30 splits. Vague language outperforms hedging in greenwashing detection based on empirical validation. The European Commission (EC) supports structured weights when theoretical priorities exist rather than equal weighting by default [64]. Effective weighting must reflect component importance systematically [25]. The ratios correspond to the intended splits: 60–40 equals 1.5 to 1 importance, while 70–30 is about 2.3 to 1.

Significant challenges exist with optimal weight combinations. Literature offers directional guidance on component importance but stops short of providing exact numerical weights. Expert consultation rounds could establish validated weight structures for broader applications, yet this falls beyond thesis research boundaries. The 60–40 and 70–30 weight allocations for electric utility analysis reflect choices grounded in available literature while recognizing the constraints of thesis research. These selections draw from theoretical evidence where possible but necessarily incorporate research limitations where empirical precision remains unavailable.

Fundamental limitations remain. The optimal weight combinations remain unknown, making sensitivity analysis necessary to test robustness [201]. Greco et al. [93] demonstrate that mixing weighting methods across analysis levels proves to effective.

2.3.6 Component Integration in Greenwashing Detection

The communication analysis requires a method to combine the five mentioned dimensions into a single greenwashing score that reflects their relative importance in theory. The PCG represents the basic mismatch between what companies actually do and what they communicate [236, 84]. Research shows this gap approach outperforms other greenwashing detection methods [15, 227]. But as discussed before: the gap alone is not enough. Additional components help the GRAT function properly. Strong evidence supports placing substantiation quality second in importance [43, 39].

Beyond these two priorities, weighting becomes a difficult task. Nemes et al. [173] could not find guidance in existing literature for how to weight different GRAT components. Communication quality dimensions lack clear theoretical hierarchies beyond substantiation quality's established role. The uncertainty about appropriate weights could undermine validity if handled poorly.

This weight uncertainty problem in greenwashing detection reflects the same challenges seen in the performance component integration in Section 2.2.5. Communication analysis faces more theoretical ambiguity than performance measurement. The greenwashing literature gives us little direction on how to weight different components, except for the priorities mentioned.

This uncertainty makes greenwashing detection well suited for the constrained ensemble approach established in Section 2.2.5. When theoretical guidance falls short, systematically exploring valid weight combinations becomes necessary. Greenwashing frameworks need transparency about methodological uncertainty rather than false precision through arbitrary weight selection [93]. The ensemble approach recognises that multiple reasonable weighting methods can exist while still maintaining theoretical constraints where evidence supports them [106]. This hybrid approach aligns with Greco et al. [93] findings established in Section 2.3.5, showing that ensemble methods effectively handle framework-level uncertainty while fixed weights manage subcomponent allocation. Guidelines from the OECD 2008 recommend uncertainty analysis for composite indicators to test robustness. Constraint-based methods handle multiple theoretical priorities, avoiding claims of precision that are not accurate [106]. Final scores combine all valid options, showing methodological uncertainty instead of hiding it.

2.3.7 Sector Considerations

Electric utilities work differently than other sectors. Their regulated status creates unique stakeholder relationships, driven primarily by compliance rules and institutional checks instead of competitive pressure [140, 49]. Long asset lifecycles create tension between short-term regulatory demands and long-term infrastructure planning. This makes future-oriented communication more structurally embedded in how utilities operate [216, 108].

Given these specific features, it is necessary to take a more specific approach regarding their environmental communication patterns. Utilities often adopt standardised, technically dense reporting formats that differ from the narrative styles seen in consumer-facing industries [112, 40]. This structural difference affects how they communicate about environmental issues and what patterns researchers should look for when analysing their reports.

2.4 Technical Framework

The multi-dimensional communication analysis framework developed in Section 2.3 creates a practical challenge: how can researchers systematically analyse five distinct communication dimensions across hundreds of corporate sustainability reports? This section establishes the theoretical foundation for computational approaches that make such analysis feasible while maintaining a scientific foundation.

2.4.1 Computational Needs

The framework established in Section 2.3 requires analysing the different dimensions simultaneously across large document sets. Traditional manual content analysis cannot handle this scale. Human coding approaches face severe scalability limits when dealing with massive modern datasets [229]. Modern research takes an "N = all approach" because computational analysis proves more cost-effective than human coding [229].

Manual coding is prone to personal bias and time constraints that compromise reproducibility and limit analytical scope [246]. Computational text analysis applies the same rules in the same way every time, eliminating inter-coder variation while creating transparent, repeatable workflows [79]. With the increasing amount of corporate sustainability reporting, manual analysis methods are no longer practical [190].

For these reasons, NLP is now widely used in corporate communication research. NLP methods make it possible to evaluate in a consistent way with consistency and level of detail in the analysis that manual coding cannot match across large document sets [190]. Computational analysis can reveal "manipulative or strategic language" patterns that are easy to miss without automation, especially for the subtle linguistic strategies identified in Section 2.3 [116]. Over 60 studies now use NLP to detect misleading climate communication, showing broad agreement among researchers that automation supports attempts to address greenwashing [251].

2.4.2 Human-AI Collaboration

Computational approaches are most effective when paired with expert human judgment. An "expert-in-the-loop" approach, combining domain expertise with computational models leads to more accurate classification of environmental reports. [104]. Automated text mining saves time although it still depends on expert judgment to set categories, check results, and interpret findings [104].

Hybrid setups pair the speed of computation with human insight, providing accuracy through close monitoring. [142]. This approach keeps researchers "in the loop" by having them verify results, fine-tune the models, and help shape the overall approach [152]. Collaboration helps capture the finer points of the topic accurately.

2.4.3 Rule-Based Methods

When analysing sector-specific corporate claims in compliance contexts, rule-based systems with clearly defined logic are often preferred over machine learning models due to their transparency and suitability for regulatory interpretation [83, 199]. Rule-based approaches support the integration of targeted knowledge specific to the sector, making their outputs easier to audit and explain [85, 83].

General-purpose machine learning models present different challenges. They may uncover spurious correlations or overfit to irrelevant patterns, while rule-based models aim to apply regulatory guidelines and keep consistent reasoning [99]. This transparency builds trust and allows for independent checks. Important when legal accountability or public scrutiny is involved [199].

Practitioners in healthcare and finance confirm this preference. Studies show that rule-based classifiers are often chosen when interpretability, verification, and sector relevance matter most [3, 212]. Domain-specific language models like ClimateBERT extend this framework further. They incorporate context-sensitive sentiment analysis tailored to climate-related communication, addressing the limitations of general sentiment tools in specialised domains [242].

It allows communication theory to be applied broadly without losing clarity and sectoral relevance. The advanced communication analysis theory developed in Section 2.3 can be implemented reliably with the transparency that regulatory contexts demand.

2.5 Tool Validation

The GRAT developed through sections 2.1–2.4 creates the need for rigorous validation. This hybrid approach requires a validation design to handle its methodological complexity.

2.5.1 Validation Approach for Hybrid Frameworks

When frameworks integrate multiple dimensions, validation becomes particularly challenging. Frameworks built from concepts that cross disciplinary boundaries require careful validation design [158]. The complexity comes from having to validate both quantitative performance measurements and qualitative communication pattern recognition.

This applies directly to greenwashing detection. The GRAT created in this research treats greenwashing as both a performance gap and a communication quality issue. Performance components require statistical testing against objective benchmarks [192]. Communication components need interpretive validation that checks whether linguistic pattern detection works [91, 208].

Documented greenwashing cases provide an external validation approach that addresses both requirements simultaneously. Real-world accusations test whether the GRAT can identify actual greenwashing behaviour while validating both performance calculations and communication pattern recognition against expert decisions. This approach tackles both measurement validity and interpretive validity through a single, robust validation method [91].

2.5.2 External Validation Through Documented Accusations

Documented greenwashing accusations from credible sources offer objective validation benchmarks for greenwashing detection [109]. These cases represent expert decisions that specific companies engaged in greenwashing. Creating “ground truth” data for testing accuracy [109].

This approach addresses a core problem in greenwashing research: assessing criterion validity despite the lack of a perfect comparison standard [1]. Documented accusations provide the most reliable proxy for identifying greenwashing objectively. Companies facing serious greenwashing accusations become a known positive control indicator. Companies with strong environmental reputations and clean records serve as negative control indicators. Whether the GRAT can distinguish between these groups provides direct evidence of discriminant validity. Statistical testing becomes necessary to determine whether observed score differences between groups reflect true discrimination rather than random variation [171]. This real-world testing connects theoretical work from sections 2.1–2.3 with practical outcomes that is important for stakeholders.

2.5.3 Limitations of Existing External Validation Approaches

Greenwashing detection frameworks face validation challenges because established measurement standards do not exist for the utility sector. ESG ratings are the most widely used sustainability metrics available, but research shows they are not suitable for validating greenwashing [123]. ESG scores capture how companies communicate their environmental efforts rather than their actual environmental impact [123].

The London Stock Exchange Group (LSEG) methodology demonstrates this problem. Their scores weight industry materiality and disclosure transparency heavily instead of verified environmental outcomes [138]. Companies with high ESG scores actually face more greenwashing accusations [123]. ESG scores correlate positively with apparent performance but negatively with real environmental performance. ESG ratings therefore cannot serve as reliable greenwashing detection benchmarks.

Academic approaches focus on disclosure-performance gaps [173] but miss the multi-dimensional communication analysis central to this research. These limitations make validation through documented real-world cases the most appropriate approach for testing greenwashing detection accuracy.

2.5.4 Internal Validation

Any combined indicator requires systematic testing to verify reliability and robustness [178, 201]. A tool such as the one in this research which combines multiple components with different weights, could make results sensitive to methodological choices. Sensitivity analysis reveals quality and ranking reliability [201]. Weight variation testing shows how component weight changes affect final scores and company rankings. When small weight adjustments cause dramatic changes in results, the GRAT lacks robustness [157]. Stable results across different weight ranges build confidence [178]. When guidance on weighting is limited in the literature, mentioned in 2.3, thorough testing becomes important to validate reliable results.

3 Methodology

This chapter turns the theoretical foundations and frameworks from chapter 2 into a practical research design for creating a tool that detects greenwashing among electric utility companies within Europe. The approach builds on Delmas and Burbano [48] PCG concept while adding the multi-dimensional communication analysis dimensions.

Figure 5 shows the six-stage research design: sample selection, data collection, performance analysis, communication analysis, integration, and validation. The process starts with companies that appear in CDP datasets and RE for 2021 and 2022. Performance analysis measures different components and combines them using constrained ensemble methodology. Communication analysis examines the dimensions through NLP methods. Both of the streams feed into integration using constrained ensemble methodology before final validation through legal cases and ESG benchmarking.

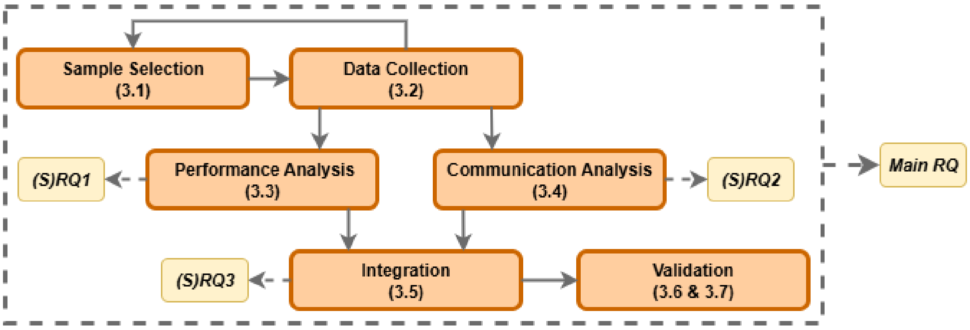

Figure 3.1: Overall Methodology Design

The methodology directly addresses each research question through specific analytical pathways. RQ1 asks *"what environmental performance variations exist among European electric utility companies based on systematic analysis of their available sustainability data?"* This question is answered through the performance analysis stage.

RQ2 asks *"what multi-dimensional communication patterns characterise environmental disclosure in European electric utility companies?"* The communication analysis addresses this by applying NLP methods based on the conceptual framework to sustainability reports.

RQ3 asks *"what greenwashing risk patterns are revealed when environmental performance and communication characteristics are analysed together?"* The integration stage addresses this by combining performance scores with communication dimensions, revealing risk patterns that neither analysis captures alone.

The main research question about *"what greenwashing risk assessment tool can be developed ..."* is answered through the complete research process. It results in a usable tool with clear application guidance and early validation that shows promising potential.

3.1 Research Design and Sample Selection

The research design and sample selection establish the foundation for systematic data collection across multiple sources.

3.1.1 Sample Selection Criteria

Sample selection is based on data availability, ensuring all parts of the GRAT are well covered. Companies must appear in CDP datasets, RE Data, and have accessible English sustainability reports. All this for both years (2021-2022). The years 2021 and 2022 were selected because CDP data for these years was available when the research began. Using an intersection approach guarantees complete data for both performance measurement and communication analysis components developed in Chapter 2. Within the CDP datasets, 43 companies list "power generation" as their primary industry and indicate a European country as their "Country/Area", reflecting the location of each company's headquarters or registered office. The 43 companies operate in countries either within the EU or closely connected to EU environmental policy frameworks as a result of regulatory harmonisation influenced by the Brussels effect [12]. Most companies operate in countries that extend beyond formal EU borders, but their regulatory connections make them relevant for European policy analysis.

Both 2021 and 2022 CDP data must be available for temporal analysis. Chapter 2.2.4 shows that tracking changes over time reveals trends and patterns that single-year data misses, improving reliability for performance assessment [78, 239]. While this requirement reduces the available sample, it supports the detection of performance shifts over consecutive years changes that are important to this component of the GRAT.

Finding external third-party data proved to be challenging. Multiple database searches revealed limited coverage for most companies when looking for consistent datasets. The RE data stream covered 27 of the 43 CDP companies with sufficient information for both years. Further screening showed that 17 companies had complete data across CDP, RE, and sustainability reports. However, three of these reported market-based scope 2 emissions instead of location-based ones in their CDP disclosures. Because the GRAT needs consistent emission methods for fair comparison, this reduced the final usable sample. Data constraints shaped the final sample selection, illustrated by the feedback loop in *Figure 5*. All chosen companies have complete and consistent information across three sources: CDP (self-reported metrics), RE third-party verified data, and sustainability reports for the full two-year period. However, this approach creates systematic bias, as sample selection was driven entirely by data availability rather than representativeness. Only companies with complete data across all three sources were included, which may introduce selection effects where companies with better data management practices also exhibit different greenwashing behaviours compared to the broader sector. The implications of this systematic bias between data management quality and environmental communication patterns are discussed further in Chapter 6.

3.1.2 Final Sample Composition

Based on the data constraints outlined in Section 3.1.1, the final sample consists of 14 electric utility companies within Europe with complete CDP, RE, and sustainability report coverage for both 2021 and 2022. The companies represent diverse business models spanning vertically integrated utilities, renewable energy specialists, and infrastructure investment companies across multiple European countries plus Turkey. See Appendix A for complete company descriptions and characteristics.

The sample shows substantial variety in company characteristics while maintaining data completeness across the core GRAT dimensions. Despite the data-constrained sample size, these 14 companies span different business model categories and geographic locations within Europe. This provides a foundation for testing the GRAT across the varied organizational contexts present within this specific sample.

3.1.3 Temporal Scope Selection

The two-year analysis period (2021-2022) addresses specific requirements for detecting greenwashing through temporal dynamics rather than single-point measurements. To clarify: the 2021–2022 period refers to the two full reporting years, covering all of 2021 and 2022. All 14 companies follow the same reporting timeframe, from January 1st to December 31st for both years. Greenwashing detection benefits substantially from examining changes over time, as companies may engage in selective disclosure and target manipulation that becomes visible through temporal analysis of corporate commitments and their modifications over time. A two-year period offers enough data to spot these patterns while keeping the analysis practical and not too complex.

Literature on temporal dynamics in greenwashing detection supports multi-year approaches for several reasons. Montgomery, Lyon, and Barg [159] show that companies use "uncertainty about the future as a vehicle for greenwashing by pledging future action that may never arrive," demonstrating the importance of tracking commitments over time. According to Jørgensen et al. [118], corporate emissions goals are often undermined by weak accountability: with "320 (31%) disappeared" targets and limited stakeholder oversight of target modifications. The timing and frequency of sustainability commitments shape how credible they seem, requiring temporal analysis to assess consistency [154]. Two years allow for trend analysis while reducing single-year anomalies that might affect the accuracy of the GRAT's results.

The 2021-2022 period also provides important regulatory context as a pre-CSRD implementation baseline. The CSRD started in 2024, making this analysis period representative of corporate environmental communication before standardised EU reporting requirements took effect. This period provides important understanding of voluntary disclosure behaviours before mandatory frameworks/rules were introduced, providing a baseline for understanding how utilities communicated environmental information without regulatory standardisation. Future research can use this period as a reference point for measuring changes in communication patterns following CSRD implementation. It is important to note that the analysis period coincides with the COVID-19 pandemic, a factor that may have shaped both corporate activity and disclosure practices. This transitional context may influence both environmental performance metrics and communication strategies, affecting the interpretability of results. Chapter 6 further addresses how this transitional period impacts the findings.

3.2 Data Collection Methods

With the research design and sample selection established, the methodology proceeds to data collection procedures across multiple sources.

3.2.1 Environmental Data Sources

Chapter 2.2.1 established the theoretical need for using multiple data sources to address the credibility paradox in environmental reporting. This section operationalises that approach by collecting performance data from both self-reported (CDP) and third-party verified RE sources, following the conceptual framework.

CDP Data Source
CDP has become the main source for climate disclosures globally [118]. The supervisor provided only the requested CDP question tabs instead of the full dataset. These targeted sections contained emissions data, target information, and financial metrics needed for the GRAT performance measure-

ment. This source delivers the self-reported performance metrics required for transparency comparison outlined in the conceptual framework.

Refinitiv Eikon Data Source
RE provides the third-party verified component of the data approach. RE offered a solution due to its reliability and transparency in academic ESG studies, as noted in Chapter 2.2.1 [214]. Environmental performance data was retrieved through the SUSFIN category in RE Datastream. This dataset includes manually verified sustainability metrics from company reports: Scope 1 and Scope 2 emissions (ton CO2e), emission intensity, reduction targets, target years, reduction percentages, and renewable energy data across multiple years. Trained analysts extract and harmonise over 400 ESG indicators, making sure results are comparable between industries without relying on modelled estimates [193].

Scope 3 Exclusion
Scope 3 emissions were excluded due to inconsistent calculation methods and data quality issues across both CDP and RE datasets. There is a risk that companies shift emissions to their supply chains rather than reducing their actual impact. Due to reporting variation, these patterns cannot be reliably compared. These limitations and their effects on greenwashing detection are discussed further in Chapter 6.

3.2.2 Sustainability Report Collection

To examine corporate environmental communication, sustainability reports provide a valuable data source. These reports have become companies' primary way to formally share environmental information with investors, regulators, and consumers [128]. While CSRD and EU Taxonomy now demand standardised, verifiable disclosures [177, 175], the 2021-2022 study period preceded these requirements. Sustainability reports during this timeframe still represented companies' main formal environmental disclosure vehicles, making them relevant for communication analysis despite a potential lack of formal report taxonomy.

All sustainability reports were downloaded as PDF directly from company websites for both 2021 and 2022 (report sources listed in Appendix B). Since CSRD taxonomy requirements had not yet taken effect during this period, companies used varied report titles including "integrated reports", "ESG performance reports", and "sustainable development reports". Selection criteria focused on identifying the main environmental disclosure document for each company-year: the longest, most complete and report containing environmental information, regardless of specific title or format. All collected reports were in English, meeting the language requirement established in section 3.1. Report length varied across companies. Document lengths ranged from more concise focused reports to extensive integrated disclosures. These structural differences reflect the pre-standardisation reporting landscape that characterises the study period, before CSRD implementation created uniform disclosure requirements.

3.2.3 Data Processing Steps

Data preprocessing transformed raw inputs from multiple sources into standardised datasets suitable for the integrated greenwashing risk analysis. This section details the cleaning and extraction procedures applied to CDP performance data, RE validation processes, and sustainability report text extraction, addressing data quality challenges identified in Chapter 2.2.1's performance measurement component of the GRAT.

3.2.3.1 CDP Data Extraction

CDP data represents the self-reported component. Even with partial data, the CDP dataset exceeded Excel's capacity and required Python-based extraction. After filtering for the 14 sample companies (explained in 3.1.1), five key data elements were extracted to capture the performance measurement components needed from this dataset.

Specific CDP disclosure sections supported the performance component of the GRAT directly. Reporting period dates (C0.2) ensured temporal alignment. Scope 1 and Scope 2 emissions in ton CO2e (C6.1 and C6.3) provided the foundation for emission intensity calculations, with location-based Scope 2 reporting verified for consistency. Scope 3 emissions were excluded due to sector-specific relevance limitations and inconsistent calculation methods across the CDP and Eikon datasets, as mentioned before. Intensity metrics (emissions relative to revenue or operation) are preferred over raw emissions because larger companies generally produce more emissions [205, 78]. The emission intensity data from the C6.10 tab addressed this issue: most companies reported revenue-based intensity measures, allowing revenue to be deducted where available. For companies using alternative intensity denominators, RE revenue data filled the remaining gaps. Reporting currency information (C0.4) enabled standardised conversion to US dollars. Currency standardisation allowed for cross-company comparability, as there is a need for standardised metrics for fair comparison [151]. Historical exchange rates from FXTOP [86] provided yearly averages for each relevant currency pair, as shown in Table 3.1.

Table 3.1: Historical conversion to US$ (source: FXTop, n.d.)

Currency	2021 Rate (to US$)	2022 Rate (to US$)
NOK	0.116	0.104
CZK	0.046	0.043
EUR	1.183	1.053
PLN	0.259	0.225
CHF	1.094	1.048
DKK	0.159	0.142

By using this approach, a consistent financial foundation was set for emission intensity and revenue-based metrics.

3.2.3.2 Refinitiv Eikon Data Extraction

Using RE data is supporting the transparency assessment outlined in Chapter 2.2.1. Access of RE Datastream was obtained by Erasmus University Rotterdam via the 'SUSFIN' category. Individual company searches led to ESG data sections, where ESG statements covering five years were downloaded as Excel files. This process captured ESG data for 27 European utilities, including all 14 sample companies plus sector comparators.

Python scripts extracted data across three years: 2020, 2021, and 2022. The 2020 data proved necessary for tracking emissions progress and detecting target changes (see section 3.3). Eight specific metrics supported different performance components:

- **Data collection period**, temporal alignment across companies
- **Auditor name**, verification context for reliability assessment
- **ESG combined score**, validation purposes for external benchmarking
- **Total Energy Use / Million in Revenue $ & Energy Use Total**, revenue calculation for companies missing CDP financial data
- **Total Renewable Energy**, renewable energy component support with intensity calculations removing unit dependency

- **Emission Reduction Target Percentage**, reduction goals for forward-looking ambition assessment
- **Emission Reduction Target Year**, target completion timelines for ambition evaluation
- **Total CO2 Emissions (ton CO2e) / Million in Revenue $**, scope 1+2 emission intensity for CDP transparency comparison detailed in section 3.3

3.2.3.3 Sustainability Report Cleaning

Extracting raw text from the sustainability reports required extra processing. Several reports used a two-column layout, which caused standard PDF-to-text extraction tools to read across the page from left to right, jumping between columns and scrambling sentences. Shen et al. [210] describe this issue and propose layout-aware extraction as a solution. Based on their approach, the reports were processed using SpaCyLayout, a layout parser designed to preserve reading order in structured PDFs.

Running SpaCyLayout came with a cost. Each report took around 30 minutes to process. To avoid repeating this step, all extracted text was saved as .txt files. The parser worked for most reports, but a few outputs included scrambled characters or symbols instead of normal letters. These encoding errors had to be corrected manually by mapping incorrect characters back to their intended form. Character replacements were followed by a few final cleanup steps. Double spaces were removed, and inconsistent spacing in chemical terms like "CO 2" was fixed to "CO2" to avoid token splitting during NLP processing (see section 3.3). Since SpaCy sees spaces as separate tokens. These fixes ensured the cleaned final .txt files were suited for further analysis.

3.3 Measuring Environmental Performance

The collected performance data now requires operational transformation into standardised environmental performance scores through systematic implementation methods.

3.3.1 Calculating Performance Scores

This section translates the performance measurement theory from the conceptual framework into practical score calculations. Each environmental performance component is converted into a numerical score between 0 and 100 using theoretically grounded formulas. This avoids normalisation within the sample, limits the influence of outliers, and keeps scores related to actual performance. The scoring follows a constrained ensemble weighting structure that systematically explores all valid weight combinations of the components while maintaining the theoretical hierarchy established in Chapter 2.2.5. An additional renewable energy bonus (up to 10 points) and a penalty for suspicious target changes (up to –10 points) are applied separately. The ensemble weighting approach reflects the theoretical priorities developed in Chapter 2.2 and will be explained in the next section (3.3.2).

3.3.1.1 Emission Intensity

Emission intensity forms the core of the performance score, as supported by ESG benchmarks and academic studies [54, 205]. To allow comparison across companies of different sizes, total Scope 1 and 2 emissions (in ton CO2e) were divided by revenue (in million US$) for each year. This avoids the bias that comes with using absolute emissions [78]. Most revenue data came from RE; when missing, values were filled in using CDP disclosures.

To evaluate relative performance, each company's emission intensity (Scope 1+2 per revenue) was benchmarked against the sector average. This sector baseline was derived from a 27-company European utility dataset, including the 14-sample companies and 13 other companies from this sector. A ratio was calculated to determine how far each company varied compared to the benchmark:

$$\text{Ratio}_{i,y} = \frac{\text{EmissionIntensity}_{i,y}}{\text{IndustryAverage}_y} \tag{3.1}$$

where:

- $\text{Ratio}_{i,y}$ = emission intensity ratio for company i in year y
- $\text{EmissionIntensity}_{i,y}$ = Scope 1+2 emissions per revenue for company i in year y
- IndustryAverage_y = sector average emission intensity in year y

Companies performing 50% better than the average (ratio ≤ 0.5) received the full 100 points. Those emitting at least twice the average (ratio ≥ 2.0) received zero points. For companies in between, a square root decay function was applied. This approach follows Becker, Robertson, and Vandenberg [14], who show that square root scaling reduces score distortion caused by outliers while keeping meaningful performance differentiation:

$$\text{Score}_{i,y} = \begin{cases} 100 & \text{if Ratio} < 0.5 \\ 0 & \text{if Ratio} \geq 2.0 \\ 100 \times \left(1 - \sqrt{\frac{\text{Ratio}-0.5}{1.5}}\right) & \text{otherwise} \end{cases} \tag{3.2}$$

where:

- $\text{Score}_{i,y}$ = emission intensity score for company i in year y (0-100)
- Ratio = emission intensity ratio from previous equation

This method ensures higher points for well-performing companies while remaining lower scores for underperformance. It reflects the theoretical guidance that outliers should not drive overall scores [181, 129], and that reward scaling should be realistic and interpretable across companies [14].

3.3.1.2 Goal Achievement

This component assesses whether companies stated targets are actually in line with lowered emissions compared to previous year emissions. It is a backward-looking measure that links targets to real progress, reflecting the accountability aspect of goal-setting [233, 185]. Actual progress was calculated as the percentage of reduction in Scope 1 and 2 emissions (in ton CO2e) compared to the previous year. Expected progress was based on the company's own reduction target, divided evenly across the years from the baseline to the target year. This gives the annual reduction rate the companies would need to follow to stay on track [200].

This creates a direct comparison between what companies claimed they would achieve and what they actually achieved. The formulas are:

$$\text{ActualProgress}_{i,y} = \left(\frac{E_{y-1} - E_y}{E_{y-1}} \right) \times 100 \tag{3.3}$$

$$\text{ExpectedProgress}_{i,y} = \frac{\text{TargetReduction}_{i,y}}{\text{TargetYear}_{i,y} - \text{BaselineYear}_{i,y}} \tag{3.4}$$

$$\text{AchievementRate}_{i,y} = \left(\frac{\text{ActualProgress}_{i,y}}{\text{ExpectedProgress}_{i,y}} \right) \times 100 \tag{3.5}$$

where:

- $\text{ActualProgress}_{i,y}$ = actual emission reduction percentage for company i in year y
- E_{y-1}, E_y = emission intensity in previous year and current year respectively
- $\text{ExpectedProgress}_{i,y}$ = expected annual reduction rate based on stated targets
- $\text{TargetReduction}_{i,y}$ = total reduction percentage target for company i in year y
- $\text{TargetYear}_{i,y}$ = target completion year for company i in year y
- $\text{BaselineYear}_{i,y}$ = baseline year for target calculation
- $\text{AchievementRate}_{i,y}$ = ratio of actual to expected progress (percentage)

Scores were based on how closely companies tracked to their target trajectory. Perfect tracking (100%) was not required for full points; companies exceeding expectations received the maximum score if their achievement rate was 200% or higher. This approach led to a more balanced distribution of scores. A company making no progress received zero. The score was scaled linearly within these limits, following recommended practices for clarity and comparability [13]:

$$\text{Score}_{i,y} = \begin{cases} 100 & \text{if AchievementRate} \geq 200 \\ 0 & \text{if AchievementRate} \leq 0 \\ 100 \times \left(\frac{\text{AchievementRate}}{200} \right) & \text{otherwise} \end{cases} \tag{3.6}$$

where:

- $\text{Score}_{i,y}$ = goal achievement score for company i in year y (0-100)
- AchievementRate = achievement rate from previous equation

By rewarding steady advancement on long-term goals, this approach reflects goal-setting principles and ensures clear understanding across different emission paths [118, 233].

3.3.1.3 Target Ambition

This component looks at how ambitious companies are in their forward-looking emissions targets. It uses the target year and the stated reduction percentage (from RE) to measure the annual improvement that the company intends from this year forward. Separating this dimension from achievement prevents rewarding companies that set low targets (in order to receive a high achievement rate) while penalising those with ambitious targets who have not (yet) reached them [115]. It also helps prevent inflated overall scores when companies set low-effort targets and then outperform them. This reduces the risk of scoring benefits from strategic under-commitment, a tactic linked to sandbagging and target manipulation [249, 115].

To calculate ambition, the reduction percentage was divided by the years left until the target year, with the current year (2021 or 2022) as the baseline. This prevents companies from appearing ambitious by setting long-term targets early and then take minimal action to follow up [249].

$$\text{TargetIntensity}_{i,y} = \frac{\text{TargetReduction}_{i,y}}{\text{TargetYear}_{i,y} - y} \tag{3.7}$$

where:

- $\text{TargetIntensity}_{i,y}$ = annual reduction intensity target for company i in year y
- $\text{TargetReduction}_{i,y}$ = total reduction percentage target for company i in year y
- $\text{TargetYear}_{i,y}$ = target completion year for company i in year y
- y = current assessment year (2021 or 2022)

As an example for $TargetYear$: a 50% reduction target by 2030 would mean a 6.25% decrease per year if stated in 2022 (y = 2022).

To convert this into a score, a linear scale was applied, giving full points to companies aiming for at least 10% annual reduction, aligned with pathways that meet science-based climate goals [206, 234]. Companies setting targets below 1% annual reduction received very low scores, and those with no forward-looking targets scored zero. The formula is:

$$\text{Score}_{i,y} = \begin{cases} 100 & \text{if TargetIntensity} \geq 10 \\ 0 & \text{if TargetIntensity} \leq 0 \\ 100 \times \left(\frac{\text{TargetIntensity}}{10}\right) & \text{otherwise} \end{cases} \tag{3.8}$$

where:

- $\text{Score}_{i,y}$ = target ambition score for company i in year y (0-100)
- TargetIntensity = annual reduction intensity from previous equation

The scaling keeps the score interpretable while encouraging companies to set goals that are both challenging and constrained by time [13, 21]. It also addresses the manipulation risks described in Chapter 2, where companies may inflate baseline years or stretch timelines to make minimal progress appear substantial [118, 249].

3.3.1.4 Transparency

This component checks how closely a company's self-reported emissions match third-party verified data. CDP and RE both provide Scope 1 and 2 emission intensities. Comparing these two sources helps uncover potential signs of selective disclosure, as suggested by Li et al. [134] and Network for Greening the Financial System [174]. Minor differences are expected, but larger gaps raise questions about reporting credibility, due to potential underreporting.

The comparison was performed using the following formula:

$$\text{Change}_{i,y} = \left(\frac{\text{CDP}_{i,y} - \text{Eikon}_{i,y}}{\text{CDP}_{i,y}} \right) \times 100 \tag{3.9}$$

where:

- $\text{Change}_{i,y}$ = percentage difference between self-reported and third-party data
- $\text{CDP}_{i,y}$ = self-reported emission intensity from CDP for company i in year y
- $\text{Eikon}_{i,y}$ = third-party verified emission intensity from Refinitiv Eikon

If the result was zero or positive, the company received full points for that year. This means they either reported more emissions to CDP than what appeared in Eikon, or their data was consistent. When the third-party value was higher, the score decreased. If Eikon emissions were 50% or more above the CDP value, the company received zero. In between, a linear scale applied:

$$\text{Score}_{i,y} = \begin{cases} 100 & \text{if Change} \geq 0 \\ 0 & \text{if Change} \leq -50 \\ 100 \times \left(\frac{\text{Change}+50}{50} \right) & \text{otherwise} \end{cases} \tag{3.10}$$

where:

- $\text{Score}_{i,y}$ = transparency score for company i in year y (0-100)
- Change = percentage difference from previous equation

This method uses the proportional penalty scoring logic as recommended by Fan and Li [75] and Plaehn [188]. It allows some flexibility for minor mismatches while still penalising cases where the reported data appears incomplete. Since transparency supports the reliability of the other performance measures, it is scored separately but with less weight, consistent with its supporting role in Chapter 2 [51].

3.3.1.5 Suspicious Target Changes Penalty

This component penalises companies that change their emissions targets in ways that may undermine trust. It captures actions such as not having targets (could also be due to data availability), lowering reduction goals, or pushing deadlines further (without increasing the reduction percentage). Changes like these may indicate greenwashing itself, particularly when companies publicly announce long-term goals but quietly change or abandon them over time [159, 118].

The method compares each company's emissions target in the current year (2021 or 2022) to the one reported in the previous year. Three types of changes are analysed: missing targets, weaker reduction goals, and later target years. Each of these patterns is penalised separately, with a combined cap of -10 points per year.

1. Missing targets
Companies that failed to report any targets in a given year received the maximum -10-point penalty. This rule applied regardless of whether they had a target the year before. Removals or omissions are treated as signs of declining commitment [38].

2. Reduction percentage decreased
If a company lowered its reduction goal compared to the previous year:

- A base penalty of -2 points was applied.
- If the reduction exceeded 25%, the company received a total of -4-point penalty for this category.
- Smaller reductions triggered proportionally smaller additional penalties, using a linear scale.

3. Target year moved later When companies postponed their target year:

- A 2-point penalty was applied as a base.
- An additional 2 points were subtracted if the new target did not come with a higher reduction percentage.
- A separate linear penalty of up to -2 points was added based on how many years the target was delayed (0.4 points per year, capped at 5 years).

This scoring approach follows the penalty-reward structure from Chapter 2.2.5, where proportional deductions/penalties are applied to account for reduced credibility [36]. The logic behind separating this category from regular achievement or ambition scores is based on the risk of manipulation. Companies may change targets to appear successful, which breaks the link between performance and intention [221, 126]. By isolating target changes, it helps protect score integrity without overstating their weight.

3.3.1.6 Renewable Energy Bonus

This bonus rewards companies with relatively high renewable energy intensity who are actively making the transition. As shown in Chapter 2, this shift is especially relevant for electric utilities, where renewables represent the most direct path to lower emissions [101, 127]. Literature supports both structural and procedural indicators for evaluating renewable performance [80], but full renewable data was not available for all companies. Since it was not part of the selection criteria, this measure was added as a separate bonus rather than a core component. This choice kept the scoring consistent while still acknowledging companies making progress [183].

The bonus is split into two parts: a benchmark comparison and a growth component. Companies could earn up to 10 points per year.

1. Industry Comparison (max 8 points)
First, each company's renewable energy intensity (per million US$ revenue) was compared to the industry average in that year. A company with more renewables than the average received a better score. The formula used:

$$\text{Ratio}_{i,y} = \frac{\text{IndustryAverage}_y}{\text{CompanyRenewable}_{i,y}} \tag{3.11}$$

where:

- $\text{Ratio}_{i,y}$ = renewable energy performance ratio for company i in year y
- IndustryAverage_y = sector average renewable energy intensity in year y
- $\text{CompanyRenewable}_{i,y}$ = company renewable energy intensity per revenue

Smaller values indicate stronger renewable performance. To reduce the influence of outliers and avoid over-rewarding companies far above the average, a square root scaling function was used:

$$\text{Score}_{i,y} = \begin{cases} 8 & \text{if Ratio} < 0.5 \\ 0 & \text{if Ratio} \geq 2.0 \\ 8 \times \left(1 - \sqrt{\frac{\text{Ratio}-0.5}{1.5}}\right) & \text{otherwise} \end{cases} \tag{3.12}$$

where:

- $\text{Score}_{i,y}$ = renewable energy benchmark score for company i in year y (0-8)
- Ratio = renewable energy performance ratio from previous equation

This follows the same logic used for emission intensity, where square root decay smooths out score differences without distorting comparisons [14].

2. Growth Bonus (max 2 points)
The second part measures change over time. Companies that increased their renewable intensity from one year to the next received additional points. A ratio of 2.0 or more (doubling year-over-year) earned 2 points. No growth received 1 point. Values between 1.0 and 2.0 were linearly scaled:

$$\text{Growth}_{i,y} = \frac{\text{Renewable}_{i,y}}{\text{Renewable}_{i,y-1}} \tag{3.13}$$

$$\text{Score}_{\text{growth}} = \begin{cases} 2 & \text{if Growth} > 2.0 \\ 1 & \text{if Growth} \leq 1.0 \\ 1 + (\text{Growth} - 1.0) & \text{otherwise} \end{cases} \tag{3.14}$$

where:

- $\text{Growth}_{i,y}$ = year-over-year renewable energy growth ratio for company i
- $\text{Renewable}_{i,y}$ = renewable energy intensity in current year y
- $\text{Renewable}_{i,y-1}$ = renewable energy intensity in previous year
- $\text{Score}_{\text{growth}}$ = renewable energy growth bonus score (0-2)

This structure encourages both strong current performance and year-on-year improvement, which is in line with scoring recommendations in the literature [237, 226].

All performance data from RE and CDP sources, along with component score results, are provided in Appendix C.

3.3.2 Constrained Ensemble Scoring

This section implements the constrained ensemble scoring approach from Chapter 2.2.5 for calculating final performance scores considering weight uncertainty due to a lack of empirical foundation in the literature. This approach, adapted with different constraint structures and computational requirements, is later used for the greenwashing score as well. Unlike Monte Carlo approaches that use random sampling to estimate outcomes through repeated trials, this scoring method systematically generates all valid weight combinations within these restrictions:

- Individual component weights range from 0.05 to 0.50
- Emission intensity weight stays highest among all components
- Goal achievement takes second-highest weight, followed by target ambition, then transparency (emission intensity > goal achievement > target ambition > transparency)
- All weights sum to 1.00

Weight combinations change in 0.01 increments to limit total possibilities. This systematic process generates 2,285 valid combinations meeting these theoretical restrictions. Each component is calculated using the methods detailed in Section 3.3.1 and scaled to a 0-100 range to maintain comparability across the ensemble approach. The method computes scores across all combinations for each organisation-year observation, then saves ensemble statistics: mean, median and standard deviation, and interquartile range (IQR).

Stability testing involves analysing score variance across different weight combinations and verifying that organisational rankings remain robust to weight changes. This procedure follows constrained multi-criteria decision analysis principles [92] and extends composite indicator sensitivity methods [201]. The constraints reflect the theoretical relationships that were found in literature; Emission intensity is central to performance evaluation, as it reflects actual outcomes and is the least prone to manipulation [54]; Goal achievement reflects the degree to which companies follow through on their stated targets and links directly to accountability [103]; Target ambition, while forward-looking, provides important insight into the credibility of commitments [115]; Transparency supports the reliability of other indicators but is not a standalone outcome, which is why it carries the lowest weight [51]. This implements OECD [178] guidance on uncertainty analysis for composite indicators and follows the theoretical foundation established in Chapter 2.2.5.

In addition to the four core components, two adjustments are applied to the ensemble scores. The renewable energy bonus adds up to +10 points for companies with strong renewable intensity and year-over-year growth. The suspicious target change penalty deducts up to -10 points for companies that remove or weaken their targets. These elements are not part of the core weighting structure. These elements operate outside the ensemble weighting structure since they address specific behaviours not fully captured by the main components. The bonus-penalty structure follows proportional scoring principles described in the literature [36]. The resulting median ensemble scores for 2021 and 2022 are used as the performance component in the PCG calculation for greenwashing detection (see 3.5.1), Complete statistics from the constrained ensemble approach can be found in Appendix C.

3.4 Analysing Corporate Communication

Environmental performance scoring provides one component of the GRAT, requiring complementary analysis of corporate communication patterns through text processing methods.

3.4.1 Text Processing Approach

Earlier in this chapter, the reasons for using sustainability reports and the text extraction method were explained. This research uses NLP to systematically analyse those texts. Manual coding lacks consistency and scalability. Using NLP makes pattern-driven analysis reproducible for multiple communication dimensions [229, 79].

SpaCy was selected as the main NLP library. It is quick, has a modular pipeline, and rule-based matching tools. The 'en_core_web_lg' model (large English model) provides accurate tokenisation, part-of-speech tagging, lemmatisation, and dependency parsing. These functions support detection of context-dependent phrases like quantified claims, vague modifiers, and future-oriented statements. The rule-based matcher allows transparent, token-level pattern extraction, which makes it suited for communication scoring in regulatory contexts [218, 217].

Cleaned .txt reports were processed through the full pipeline. Tokenisation breaks text into individual units (tokens), lemmatisation reduces words to their base forms, and dependency parsing stores grammatical relationships.

ClimateBERT was used selectively for environmental sentiment analysis. General models often misclassify sustainability text. ClimateBERT improves classification accuracy for climate-related topics [242]. Still, the core analysis remains rule-based to maintain interpretability and relevant to the sample, as recommended by Freitas [83] and Rudin [199].

3.4.2 Human-AI Collaboration Approach

Analysing communication patterns across five dimensions creates substantial analytical challenges. Manual pattern detection often misses subtle linguistic variations, making LLMs useful for developing comprehensive word lists and identifying rule patterns. Human expertise validates these suggestions and ensures regulatory relevance. This combination provides both pattern coverage and interpretability.

This project followed an expert-in-the-loop approach to develop the communication scoring system [104]. Claude Sonnet 4 was used in the early stages to sketch out rule logic and build initial word lists. These drafts were then revised by hand, tested in sample reports, and adjusted as needed. Claude Sonnet 3.5, run through VS Code Copilot, supported implementation and debugging. Compared to GPT-based tools, Claude 3.5 performed more consistently in data science-related coding tasks, with success rates above 60% [172].

Word lists for vague terms, hedge words, and environmental expressions were built through a mix of AI suggestion and manual review. Gaps were filled by manually checking verb, noun, and adjective patterns directly from the reports. This method is not fixed, lists can be refined further as new patterns or words appear, and the structure allows for ongoing updates over time.

Each rule was checked across different companies and years to avoid overfitting to one reporting style. Where detection was unclear or inconsistent, results were reviewed manually. Appendix D outlines the verification of each model and the steps and variables used to create the models. Additionally, Claude Sonnet 4 was used to refine certain portions of the written text for clarity and academic presentation, although all methodology development, analysis, and conclusions remain original work.

3.4.3 Multi-Dimensional NLP Implementation

Chapter 2.3.4 identified five core communication dimensions that reflect different strategies that companies use to shape environmental perception: Green Communication Intensity (which forms the communication side of the PCG analysis), Substantiation Weakness, Language Vagueness, Temporal Orientation, and Reporting Consistency. The dimensions are renamed to reflect problematic communication patterns, where higher scores indicate greater greenwashing risk. Each dimension is translated into NLP logic through standalone code modules (scripts), each targeting specific linguistic features in the reports. These modules reflect different communication strategies and are listed below:

- **Green Terms:** Extracts and counts environmentally relevant words to quantify communication volume and vocabulary variety.
- **Green Term Context Classification:** Assesses each green term by its temporal scope, quan-

tification status, and whether it reflects evidence or aspiration.

- **Sentiment Analysis:** Assigns sentiment scores to green-related sentences to assess the (overall) tone of environmental messaging.
- **Vague and Hedge Words:** Detects vague terms and hedge words to evaluate clarity and certainty in language.
- **Document Similarity:** Compares report texts across the two years to identify repeated content and assess reporting consistency.

These code modules produce multiple output variables that correspond to one of the earlier mentioned communication dimensions. All these variables generated by the different communication modules are listed in Table 3.2. These outputs form the input for the final greenwashing score, as described in Section 3.5.1. There, each variable's role and relative weight is linked back to the theoretical foundations outlined in Chapter 2. For detailed explanations, including how each variable is constructed and validated, see Appendix D. The code modules can be found in GitHub repository, see Appendix K for structure and reference.

Table 3.2: Output variables per module and connection to defined communication dimensions

Output Variable	Description	Module	Dimension
Green Term Frequency	Share of environmental words relative to all content words (excluding stop words, punctuation and whitespace)	Green Terms	Green Communication Intensity
Vocabulary Diversity	Number of unique environmental terms used relative to total unique words	Green Terms	Green Communication Intensity
Future Orientation Ratio	Ratio of future-oriented claims to past and present references	Context Classification	Temporal Orientation
Quantified Claim Intensity	Share of environmental claims containing specific quantities	Context Classification	Substantiation Weakness
Evidence-Based Claim Intensity	Share of claims referencing concrete actions or implemented measures	Context Classification	Substantiation Weakness
Aspirational Claim Intensity	Share of forward-looking commitments lacking measurable details	Context Classification	Substantiation Weakness
Average Environmental Sentiment	Overall sentiment of green-related sentences	Sentiment Analysis	Green Communication Intensity
Renewable Energy Sentiment	Sentiment expressed in renewable-related sentences	Sentiment Analysis	Green Communication Intensity
Climate and Emissions Sentiment	Sentiment expressed in climate- and emission-related sections	Sentiment Analysis	Green Communication Intensity
Vague Language Intensity	Share of sentences using vague terms such as "eco-friendly" or "sustainable"	Vague and Hedge Words	Language Vagueness
Hedge Language Intensity	Share of sentences using cautious or uncertain terms like "might" or "aim to"	Vague and Hedge Words	Language Vagueness
Timeline Specificity	Measures whether vague or conditional commitments are tied to concrete deadlines	Vague and Hedge Words	Temporal Orientation
Cross-Year Similarity Score	Combined score from TF-IDF, Jaccard, and SpaCy similarity to detect repeated content	Document Similarity	Reporting Consistency
High-Similarity Sentence Ratio	Share of sentences nearly identical between 2021 and 2022 reports	Document Similarity	Reporting Consistency

3.5 Integrated Greenwashing Analysis

The separate performance and communication analyses are integrated through systematic scoring methods to produce unified greenwashing risk assessments. *Figure 6* shows how the five NLP modules generate output variables that feed into the five communication dimensions (as also outlined in Table 3.2).

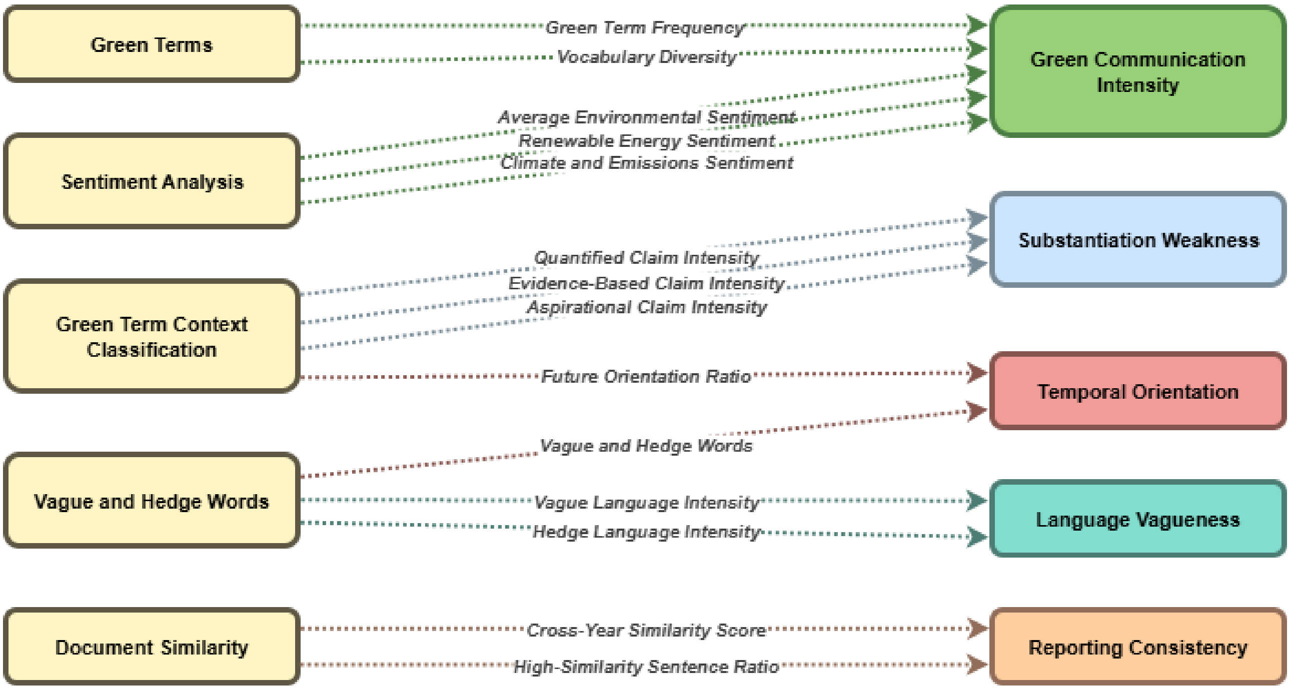

Figure 3.2: Communication Variables and Dimensions

These variables together with the performance score provide the foundation for the integrated greenwashing risk assessment detailed in the following sections.

3.5.1 Component Integration and Weighting

This section shows how individual performance scores combine with communication dimensions to build the final GRAT. The conceptual framework from Chapter 2 guides most weighting decisions. Where gaps in the literature exist, logical reasoning paired with elements from the conceptual framework determines weight allocation.

An ensemble methodology like the one in section 3.5.2 could address uncertainty in these component-level weights. However, testing all possible weight combinations across multiple nested components would create too many variations. The computational complexity would make accurate computation impossible. Instead, sensitivity analysis in Chapter 3.6 tests how robust these weighting choices are.

3.5.1.1 Output Variables

The integration process creates several key variables, as seen in *Figure 6*, that capture different aspects of greenwashing risk. This multi-dimensional approach tackles the issue that existing approaches oversimplify greenwashing as a single phenomenon when it's actually multi-faceted [173]. As research shows, certain communication patterns serve as independent greenwashing indicators regardless of performance levels [76, 231, 117]. Best practices for composite indicators recommend avoiding single-indicator dominance to prevent misleading scores [93]. The output variables, their dimensions, and justifications are presented in *Table E* (in Appendix E).

The following section outlines the calculations for each dimension applied in the final greenwashing score (3.5.2), transforming output variables into weighted, measurable components.

3.5.1.2 Green Communication Intensity Calculations

Multiple variables combine to form the green communication intensity dimension (see *Figure 6*), which covers the communication side of the PCG. Environmental Communication Theory identifies two core elements: sentiment framing and green communication volume [2, 248]. The first component is the sentiment score, which captures how companies discuss environmental issues positively as this directly affects stock prices and correlates with ESG ratings [16, 250]:

$$\mathrm{CSS}_{i,y} = 0.6 \times \mathrm{AES}_{i,y} + 0.2 \times \mathrm{RES}_{i,y} + 0.2 \times \mathrm{CES}_{i,y} \tag{3.15}$$

where:

- $\mathrm{CSS}_{i,y}$ = Combined Sentiment Score: weighted sentiment score for company i in year y
- $\mathrm{AES}_{i,y}$ = Average Environmental Sentiment: overall sentiment of green-related sentences
- $\mathrm{RES}_{i,y}$ = Renewable Energy Sentiment: sentiment expressed in renewable energy contexts
- $\mathrm{CES}_{i,y}$ = Climate and Emissions Sentiment: sentiment in climate and emissions discussions

Environmental sentiment receives the largest weight (60%) because it covers the broadest range of environmental topics. Renewable energy and climate emissions each get 20% since they represent specific domains within environmental communication that carry distinctly positive sentiment that shapes public perception [135]. This allocation prevents equal weighting while acknowledging their different conceptual scope.

The second component measures green term frequency and vocabulary diversity, based on selective disclosure theory where companies reveal positive environmental information while hiding negative aspects [140]:

$$\mathrm{CGTS}_{i,y} = 0.7 \times \mathrm{GTF}_{i,y} + 0.3 \times \mathrm{VD}_{i,y} \tag{3.16}$$

where:

- $\mathrm{CGTS}_{i,y}$ = Combined Green Term Score: weighted green communication volume for company i in year y
- $\mathrm{GTF}_{i,y}$ = Green Term Frequency: share of environmental words relative to all content words
- $\mathrm{VD}_{i,y}$ = Vocabulary Diversity: number of unique environmental terms relative to total unique words

Term frequency gets 70% weight because contextual frequency predicts information processing better than vocabulary diversity [35]. Research shows that disclosure credibility increases through frequency and consistency rather than vocabulary diversity [35]. Vocabulary diversity still has importance, since it can shape how stakeholders interpret environmental disclosures by indicating the effort put into communication [180, 119].

These components combined form the overall green communication score:

$$\text{GCS}_{i,y} = 0.4 \times \text{CGTS}_{i,y} + 0.6 \times \text{CSS}_{i,y} \tag{3.17}$$

where:

- $\text{GCS}_{i,y}$ = Green Communication Score: overall green communication intensity for company i in year y
- $\text{CGTS}_{i,y}$ = Combined Green Term Score: from Formula 13
- $\text{CSS}_{i,y}$ = Combined Sentiment Score: from Formula 12

Sentiment receives higher priority because evidence shows that sentiment has greater public impact [2, 248]. Several empirical studies support this hierarchy [30, 105].

Performance-Communication Gap Calculations
The main greenwashing risk score captures the PCG that defines classic greenwashing [236, 84]:

$$\text{GRS}_{i,y} = \text{GCS}_{i,y} - \text{PS}_{i,y} \tag{3.18}$$

where:

- $\text{GRS}_{i,y}$ = Greenwashing Risk Score: basic performance-communication gap for company i in year y
- $\text{GCS}_{i,y}$ = Green Communication Score: from Formula 14
- $\text{PS}_{i,y}$ = Performance Score: environmental performance score from Section 3.3

When companies dramatically increase their environmental communication while their performance stays poor, this reflects greenwashing in its most problematic form [5, 141]. Problematic greenwashing is defined by companies displaying strong environmental communication paired with poor performance. An amplification factor is applied to score such companies higher, better reflecting true greenwashing. This follows best practices where analysts often apply moderate amplification factors [93]:

$$\text{AS}_{i,y} = \begin{cases} 1.5 \times \text{GRS}_{i,y} & \text{if } \text{GCS}_{i,y} > \text{Med}_\text{y} \\ & \text{and } \text{PS}_{i,y} < \text{Med}_\text{y} \\ \text{GRS}_{i,y} & \text{otherwise} \end{cases} \tag{3.19}$$

where:

- $\text{AS}_{i,y}$ = Amplified Score: amplified greenwashing risk score for company i in year y
- $\text{GRS}_{i,y}$ = Greenwashing Risk Score: from Formula 15
- $\text{GCS}_{i,y}$ = Green Communication Score: from Formula 14
- $\text{PS}_{i,y}$ = Performance Score: from Section 3.3

- Med_y = Median: median score across all companies in year y

This scaling exposes imbalances by increasing relative variability [144]. Research shows this gap approach works better for detecting greenwashing than alternative methods, with empirical evidence demonstrating superior performance [15, 227].

3.5.1.3 Communication Quality Dimensions

The additional communication quality dimensions in the next sections capture different aspects of how companies might mislead stakeholders through their environmental disclosures [149].

Substantiation Weakness Calculations
This dimension builds on Clarkson et al. [43] distinction between *"hard"* disclosure (numbers, evidence, verifiable facts) and *"soft"* disclosure (aspirations, promises, unverifiable statements). Their research finds that hard disclosure is more credible than soft disclosure. Studies show worse environmental performers exhibit significantly more 'optimism' and less 'certainty' in their language [39]:

$$\text{SW}_{i,y} = 0.30 \times (100 - \text{QCI}_{i,y}) + 0.35 \times (100 - \text{EBCI}_{i,y}) + 0.35 \times \text{ACI}_{i,y} \tag{3.20}$$

where:

- $\text{SW}_{i,y}$ = Substantiation Weakness: substantiation weakness score for company i in year y
- $\text{QCI}_{i,y}$ = Quantified Claim Intensity: share of environmental claims containing specific quantities
- $\text{EBCI}_{i,y}$ = Evidence-Based Claim Intensity: share of claims referencing concrete actions
- $\text{ACI}_{i,y}$ = Aspirational Claim Intensity: share of forward-looking commitments lacking details

Evidence-based and aspirational claims receive equal weight (35%) because both relate directly to the hard versus soft disclosure distinction. Regulatory frameworks support this hierarchy: the EU Green Claims Directive requires robust, science-based evidence over aspirational claims [70]. The Federal Trade Commission demands competent and reliable scientific evidence for environmental claims [76].

Quantified claims get slightly less weight (30%) despite the fact that quantitative disclosures are associated with greater clarity of information and investor trust [156]. Numbers can be manipulated or presented selectively, which explains this lower weighting.

Language Vagueness Calculations
This dimension separates vague language that creates greenwashing risk from legitimate hedging that expresses appropriate uncertainty. Research shows that vague language is a form of greenwashing that creates scepticism and damages authenticity [231, 84, 252, 203]:

$$\text{LV}_{i,y} = 0.70 \times \text{VLI}_{i,y} + 0.30 \times (100 - \text{HLI}_{i,y}) \tag{3.21}$$

where:

- $\text{LV}_{i,y}$ = Language Vagueness: language vagueness score for company i in year y
- $\text{VLI}_{i,y}$ = Vague Language Intensity: share of sentences using vague terms (e.g., "eco-friendly")

- $\text{HLI}_{i,y}$ = Hedge Language Intensity: share of sentences using cautious terms (e.g., "might")

Vague language gets 70% weight because vague wording such as *'eco-friendly'* and *'green'* means nothing concrete [209, 26]. This type of language reduces consumer trust regardless of actual performance [57, 117].

Legitimate hedging receives different treatment. It expresses appropriate uncertainty, which is essential for credible scientific communication [235, 111] and correlates with credibility expectations [136]. When companies choose words carefully through hedging, they build stakeholder trust [153, 186, 204]. The 70-30 split reflects empirical validation where distinguishing vague language from legitimate hedging achieved 86.34% accuracy in detecting greenwashing [32].

Temporal Orientation Calculations
This dimension identifies the pattern of companies making distant future promises without near-term action plans. Montgomery, Lyon, and Barg [159] found companies making impressive net-zero promises for 2050 while not making any recent progress. The United Nations [231] points out the risk of misleading stakeholders when short-term plans are absent:

$$\text{TO}_{i,y} = 0.60 \times \text{FOR}_{i,y} + 0.40 \times (100 - \text{TS}_{i,y}) \tag{3.22}$$

where:

- $\text{TO}_{i,y}$ = Temporal Orientation: temporal orientation score for company i in year y
- $\text{FOR}_{i,y}$ = Future Orientation Ratio: ratio of future-oriented claims to past and present references
- $\text{TS}_{i,y}$ = Timeline Specificity: measure of concrete deadlines tied to commitments

Future orientation gets 60% weight because without near-term goals, long-term commitments lack credibility [102]. Real commitment requires long-term pledges paired with detailed short-term actions and milestone tracking [182, 230]. When companies fail to meet environmental promises, this "links directly to greenwashing and legitimacy loss" [100, 225].

Research on temporal framing shows companies that separate future and present communication tend to invest less in future-focused activities like sustainability [137]. Meanwhile, environmental messages that reference the past rather than the future boost both credibility and people's willingness to act [11]. Timeline specificity stays important since hopeful messages need specific timelines instead of vague future promises [110].

Reporting Consistency Calculations
This dimension detects symbolic rather than substantive reporting through repetitive language patterns. Choi, Chung, and Young [41] found companies systematically repeat structure and content across years, suggesting symbolic rather than substantive reporting:

$$\text{RC}_{i,y} = 0.70 \times \text{CYSS}_{i,y} + 0.30 \times \text{HSSR}_{i,y} \tag{3.23}$$

where:

- $\text{RC}_{i,y}$ = Reporting Consistency: reporting consistency score for company i in year y

- $\text{CYSS}_{i,y}$ = Cross-Year Similarity Score: combined TF-IDF, Jaccard, and SpaCy similarity score
- $\text{HSSR}_{i,y}$ = High-Similarity Sentence Ratio: share of sentences nearly identical between years

Cross-year similarity gets 70% weight because analysis of over 10,000 sustainability reports found widespread boilerplate language that correlated with ESG rating problems [240]. Half of US companies use generic, repetitive language in ESG filings [42]. While audits check facts, they ignore how communication quality develops over time [107], which allows symbolic compliance.

Research in NLP supports distinguishing between cross-year and sentence-level analysis. Studies show sentence similarity methods using Sentence-BERT models work much better than keyword matching for detecting actual content changes [120]. Their analysis revealed monotonous thematic patterns in sustainability reports. Sentence-level analysis provides evidence of template use but plays a secondary role to thematic repetition patterns.

Weight Allocation and Normalisation
The structured weight patterns here follow evidence-based hierarchies from existing research. When studies indicate a 'clear but moderate preference', 0.6-0.4 splits represent this hierarchy while preserving the role of secondary components [64]. Stronger empirical evidence supports 0.7-0.3 splits. This approach aligns with the principle that effective weighting should systematically reflect component importance [25]. Acknowledging the uncertainty of optimal weights, sensitivity analysis becomes necessary to test the GRAT's robustness [201]. Research shows that mixing weighting methods across analysis levels proves effective [93].

Each dimension mentioned is scaled from 0 to 100 using min-max normalisation. This approach, commonly used to correct for variation in document length and linguistic style, supports fair comparison and integration into the final greenwashing score in Section 3.5.2. This normalisation approach supports fair cross-component comparisons and allow for methodological consistency [21]. By standardising component scales, the risk of bias from varying measurement ranges is reduced, increasing robustness of the created GRAT.

3.5.2 Final Score Calculation

This section applies the constrained ensemble scoring methodology established in Section 2.3.6 and operationalised in Section 3.3.2 to final greenwashing risk score calculation. The approach addresses weight uncertainty in greenwashing detection due to the lack of empirical foundation in framework literature identified by Nemes et al. [173]. Following the same systematic enumeration approach, this method generates all valid weight combinations within these greenwashing-specific restrictions:

- Communication dimension weights range from 0.05 to 0.5
- PCG weight remain the highest among all individual dimensions
- Substantiation weakness takes second-highest weight
- All weights sum to 1.00
- Weight combinations change in 0.01 increments to limit total possibilities

This systematic process generates 59,881 valid combinations meeting these theoretical restrictions (see Appendix H for statistical results). The large increase from 2,285 performance combinations occurs because performance scoring requires four ordered relationships while greenwashing scoring needs only two fixed priorities. The remaining three dimensions can vary freely within their weight ranges, creating significantly more valid combinations.

The technical implementation differs from performance ensemble scoring in two ways. First, the hybrid approach becomes necessary here due to computational limits. Testing all possible subcomponent weights within communication dimensions would create millions of combinations, making systematic enumeration impossible. Second, no bonus-penalty adjustments apply since all greenwashing risk factors integrate directly into the five-dimension structure. Extending ensemble methodology to subcomponent weights within communication dimensions would create computationally prohibitive combinations, making the hybrid approach necessary where system-level uncertainty uses ensemble methods while subcomponent allocation relies on literature-based fixed weights [93]. The method computes scores across all combinations for each organisation-year observation, then again saves ensemble statistics: mean, median and standard deviation, and IQR.

The constraints reflect the theoretical relationships that were found in literature: the PCG measures the fundamental disconnect between company actions and communications [236, 84], while Substantiation Weakness ranks second, proven in literature as the key component following the PCG [43, 39]. By fully enumerating feasible weights within theoretical limits, this approach solves the uncertainty issue Nemes et al. [173] avoiding arbitrary weight choices.

3.6 Validation Methodology

The GRAT relies on two types of validation to assess reliability and accuracy. Internal validation tests methodological robustness and reliability through multiple approaches applied throughout this research: sensitivity analysis of weight variations (detailed below), verification of NLP modules in Appendix D, and ensemble methods that reduce the risk of arbitrary weight selection by grounding all choices in literature. External validation examines real-world accuracy using documented greenwashing cases to test whether the GRAT correctly identifies actual greenwashing behaviour, though with significant statistical limitations due to the small sample size that limit conclusions to indicative findings only.

3.6.1 Sensitivity Analysis and Robustness Testing

The GRAT requires internal validation, testing robustness through systematic weight variation testing across communication components. Following Saisana, Saltelli, and Tarantola [201], sensitivity analysis reveals GRAT quality by measuring how component weight changes affect final scores and company rankings. Small weight adjustments that cause large ranking changes indicate poor robustness [201]. The analysis addresses theoretical uncertainty identified in Chapter 2.3.5 about optimal sub-component weights within communication dimensions.

The ensemble method in Section 3.5.2 handles uncertainty across the five main dimensions, while fixed weights for sub-components within each dimension are based on literature. These sub-component weight allocations need robustness verification since guidance on actual numerical weighting remains limited in greenwashing literature [173]. To test this robustness systematically, the analysis examines how sub-component weight variations affect both individual scores and company rankings.

3.6.1.1 Scenario Development

Scenario development introduces controlled changes in sub-component weights to assess their effect on final scores and company rankings. This helps determine whether the GRAT produces stable results or if minor methodological adjustments lead to relatively large shifts in outcomes. The analysis tests 17 systematic scenarios: the baseline using literature fixed weights and 16 variations applying weight adjustments to each sub-component, as shown in *Table D1* in Appendix D. Most components receive $\pm 10\%$ variations while substantiation weakness receives $\pm 5\%$ adjustments to preserve the theoretical hierarchy Evidence $\geq$ Aspirational $>$ Quantified [43]. All weight modifications still sum to 1 while respecting literature-based component priorities.

Each scenario recalculates all communication dimension scores with modified weights, then processes results through the complete ensemble methodology established in Section 3.5.2. This approach tests both sub-component sensitivity (Appendix I) and its interaction with system-level weight uncertainty (see Appendix H). Every scenario runs through all 59,881 valid ensemble weight combinations, keeping the methodology consistent with the main final greenwashing risk analysis.

3.6.1.2 Statistical Analysis

Following Saisana, Saltelli, and Tarantola [201], the analysis measures ranking stability using $\bar{R}_S$ (average ranking shift) across scenarios compared to baseline rankings. $\bar{R}_S$ values below 2.0 indicate high robustness, 2.0-4.0 suggests moderate stability, while values above 4.0 signal low robustness requiring methodological attention.

Beyond ranking analysis, absolute score sensitivity uses three metrics [93, 178]. Mean Absolute Deviation (MAD) measures average score shifts from baseline values. The Coefficient of Variation (CV) provides standardised variability measures across different score levels, while Score Range captures the maximum uncertainty span for each company. Companies with CV values above 15% or score ranges exceeding 10 points are classified as highly sensitive to methodological choices.

The analysis ranks scenarios by their $\bar{R}_S$ values and average MAD across all companies to identify which sub-component weight changes cause the largest disruptions. This approach addresses both stakeholder concerns about relative company positions and policy needs for reliable absolute greenwashing risk assessments. The results separate components that primarily influence rankings from those that affect the actual score values. The analysis identifies companies whose assessments remain stable across all scenarios versus those requiring careful interpretation. Detailed sensitivity testing results appear in Appendix I, with complete statistical outputs.

3.6.2 Known Case Validation

While sensitivity analysis confirms internal methodological reliability, external validation depends on testing the GRAT against real-world greenwashing cases to evaluate how well the GRAT performs in practice. However, with only 14 companies in the sample, this validation has serious statistical power problems because the central limit theorem requires larger samples to work properly. The validation results should be seen as indicative only, not as strong statistical evidence. Having more companies would give much better results, but this represents all the data we could access given the constraints. This external validation is done by performing an online search examining all 14 sample companies between January 2020 and December 2023, identifying accusations documented by credible sources. These documented accusations serve as validation benchmarks, creating 'ground truth' data for testing GRAT accuracy [109].

The approach used multi-language search terms since local media and Non-Governmental Organization (NGO) campaigns often get missed by English-only searches. Sources included Greenpeace national offices, investigative outlets, consumer protection agencies, and energy sector publications. Complete search methodology and results can be found in Appendix J.

Based on the search results the companies were classified into three validation categories: known positive cases (documented greenwashing accusations), clean record companies (no accusations found), and historical reference cases (limited evidence within timeframe). Because of the timeframe used in this study, past greenwashing cases are considered irrelevant and classified among the clean record group. Statistical testing compares GRAT scores between known positive cases and clean record companies to determine whether companies with documented greenwashing accusations score significantly higher than those without.

Statistical validation uses Mann-Whitney U (MWU) tests [243, 143] to compare average GRAT scores between known positive cases and companies with a clean greenwashing record (within the 2020-2023 timeframe), although this test has very low statistical power with only 14 observations. The median greenwashing scores produced by the ensemble method were averaged to reflect consistent patterns rather than yearly fluctuations. While the non-parametric MWU test works with small samples, having just 14 observations creates major limitations. The patterns we see might show genuine GRAT accuracy, or they could be random events that happened between 2021-2022. Effect size calculations help interpret the results, but the small sample size restricts statistical power and makes strong conclusions impossible [171]. A more detailed discussion of these validation limitations and their implications appears in Chapter 6.

4 Results

Drawing on the six-stage methodology from Chapter 3, this chapter presents the core empirical findings. The analysis follows three main steps: assessing environmental performance, examining communication patterns, and evaluating greenwashing risk. Each section contributes to answering the research questions, which will be addressed in Chapter 7 Conclusion. These results reflect the 2021-2022 period, a transitional time marked by COVID-19 impacts and pre-CSRD reporting practices. Since sample selection was driven by data availability rather than representativeness, this creates systematic bias as discussed in Chapter 3. The small sample size means all findings are exploratory and cannot be generalized beyond these specific companies. Readers should keep these limitations in mind when interpreting the figures and results presented below, as the implications and broader context will be discussed in Chapter 6.

4.1 Environmental Performance

This section addresses RQ1 by examining environmental performance variations across the 14 European utilities in this sample using the performance component of the GRAT established in Chapter 3.3. The exploratory analysis suggests substantial performance differences across companies, with performance scores ranging from 15.0 to 95.0 points and significant year-over-year changes spanning from -35.1 to +63.0 points.

4.1.1 Performance Components Contributions

Figure 4.1 demonstrates how the four performance components contribute differently to overall environmental achievement, with important methodological considerations affecting interpretation. The renewable energy bonus and suspicious target penalties are shown as green and red arrows, respectively.

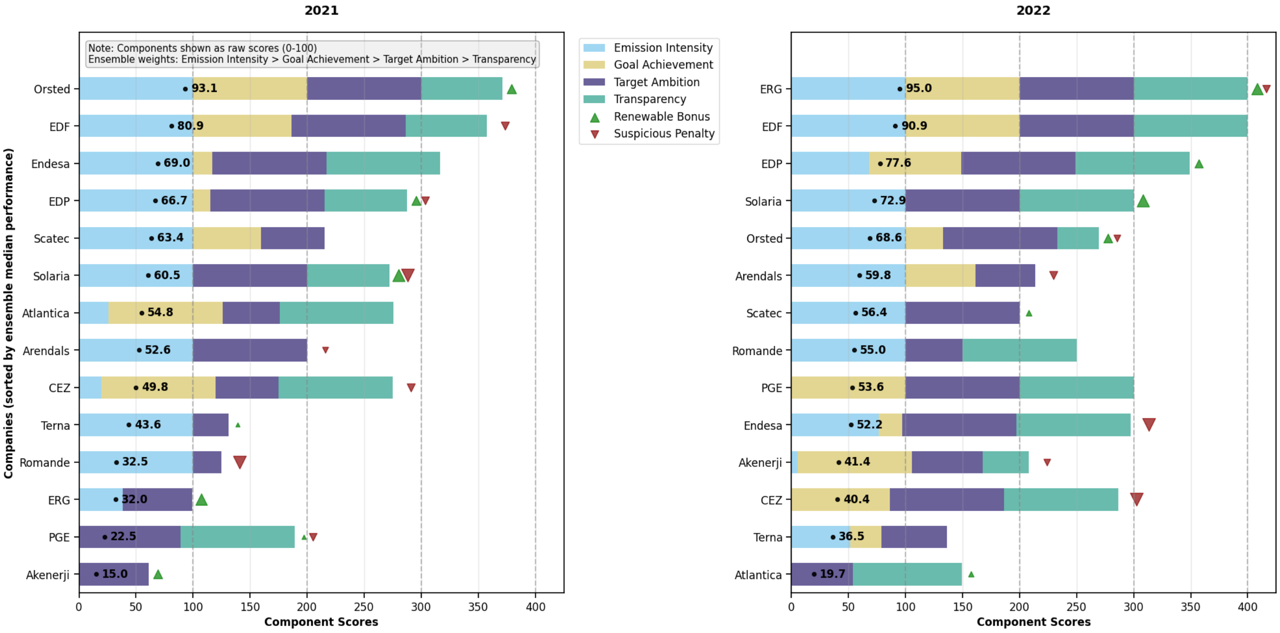

Figure 4.1: Performance Component Analysis

Emission intensity scores in *Figure 4.1* show the widest distribution, ranging from perfect scores (100.0) for companies like Ørsted and EDF to minimal scores for PGE and Akenerji (0.0-5.43). However, substantial middle-ground performance is visible among several companies, including Endesa (69.0), EDP (66.7), and Scatec (63.4). See Appendix C for the complete component scores and raw performance data used for the analysis. This distribution reflects the threshold-based scoring method, which is intended to limit outlier influence. Companies performing well above or below industry benchmarks receive maximum or minimum scores, while others are scored proportionally between those points.

The disconnect between component stack width and ensemble median scores (black dots) throughout *Figure 4.1* illustrates the constraint-based weighting methodology. While raw component scores appear as 0-100 values, the ensemble approach applies 2,285 weight combinations within theoretical constraints (emission intensity > goal achievement > target ambition > transparency). Median scores are drawn from the combinations to reflect ranking-based relationships instead of adding scores directly. This balances uncertainty in weighting while keeping the tool's priorities intact. The ensemble uncertainty ranges from minimal variation (0.1 standard deviation for top performers like EDF) to moderate uncertainty (5.82 for transitional cases like PGE 2021), demonstrating methodological stability across all different weight combinations (see Appendix C3 for complete ensemble statistics).

Goal achievement results that show perfect achievement scores (100.0) occur across very different company profiles, from clear transition leaders like EDF to coal-heavy companies such as PGE in 2022. Companies such as Arendals (2021), ERG (2021), and Solaria (both years) scored zero on goal achievement. Target ambition suggests a clear sample-wide shift toward higher scores, with most companies falling between 50.0 and 100.0 (see *Figure 4.1*). Ten companies reached the maximum ambition score across various years. Transparency scores vary widely, reflecting disclosure consistency between CDP self-reporting and RE third-party verification. Component interactions visible in *Figure 4.1* suggest that high emission intensity performance does not guarantee high goal achievement scores, indicating potential disconnects between structural performance (emission levels) and dynamic performance (goal achievement). These component patterns combine to create overall performance trajectories shown in the following analysis.

4.1.2 Performance Scores and Trajectories

The performance scores used in this research are shown as ensemble medians in *Figure 4.1*. They form a structured range of scores, from EDF's leading scores (80.91 and 90.91) to Atlantica's decline from 54.82 to 19.70. *Figure 4.2* shows that year-over-year changes tend to follow certain patterns instead of minor fluctuations.

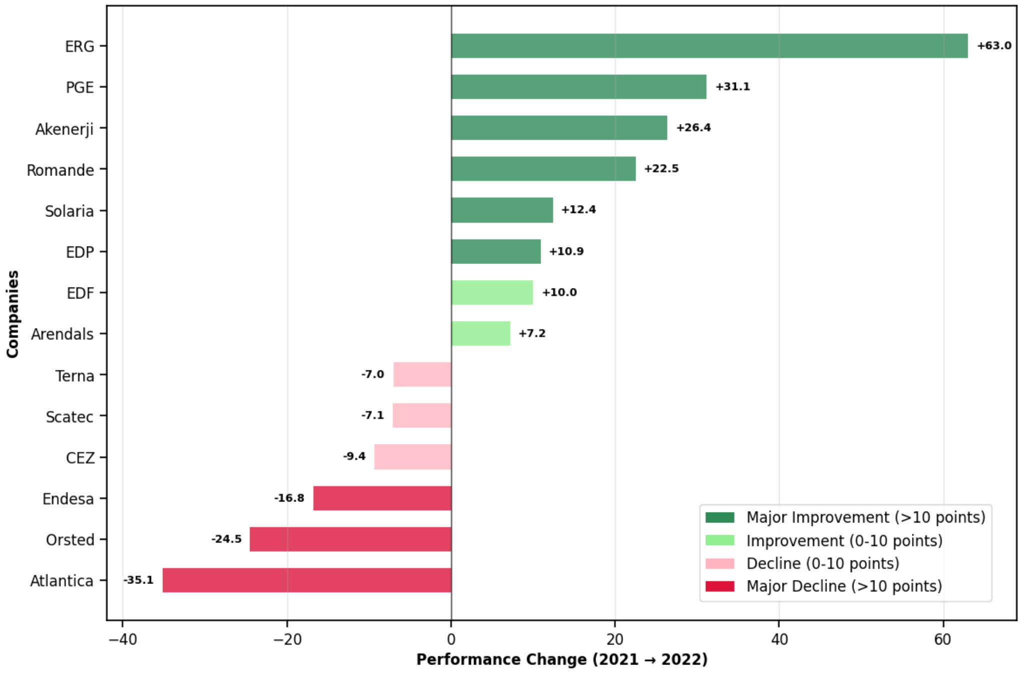

Figure 4.2: Year-over-year Performance change

The most notable shift in the sample comes from ERG, which increased from 32.04 to 95.04, supported by perfect component scores visible in *Figure 4.1*.

PGE's +31.1 improvement presents a different pattern: achieving a 100 score for goal achievement by 2022 despite maintaining relatively high emissions. The change in score is mostly due to large decrease in reported emissions intensity, which drives the goal achievement component.

Atlantica's -35.1-point drop reflects weaker performance in both emission intensity and goal achievement, even though its target ambition remained unchanged. Ørsted's -24.5-point decline occurs despite relatively strong component scores, with a slight drop in goal achievement and transparency contributing to the change.

Figures 4.1 and *4.2* suggest clear differences in company performance, which would be expected to correspond with varied communication approaches. High performers such as EDF and ERG have strong environmental results to support their claims, while companies like PGE and Atlantica must communicate their efforts despite weaker performance metrics. These performance realities create the foundation for examining how utilities actually communicate about their environmental initiatives. The question becomes whether communication intensity and quality align with demonstrated performance capabilities, or whether inconsistencies arise that could indicate potential greenwashing.

4.2 Multi-Dimensional Communication Analysis

This section addresses RQ2 through systematic analysis of communication patterns in sustainability reports using the multi-dimensional NLP component of the GRAT. The exploratory analysis suggests distinct communication strategies across five dimensions, with companies showing varied patterns regardless of their environmental performance levels.

4.2.1 Communication Intensity Patterns

The GRAT communication component shows variation across five dimensions; these operate independently of performance levels (except for green communication intensity). *Figure 4.3* shows that green communication intensity is based on a weighted combination of environmental term frequency and sentiment analysis. Both components are normalised using min-max scaling, where zero reflects the lowest value in the sample, not a complete absence of communication.

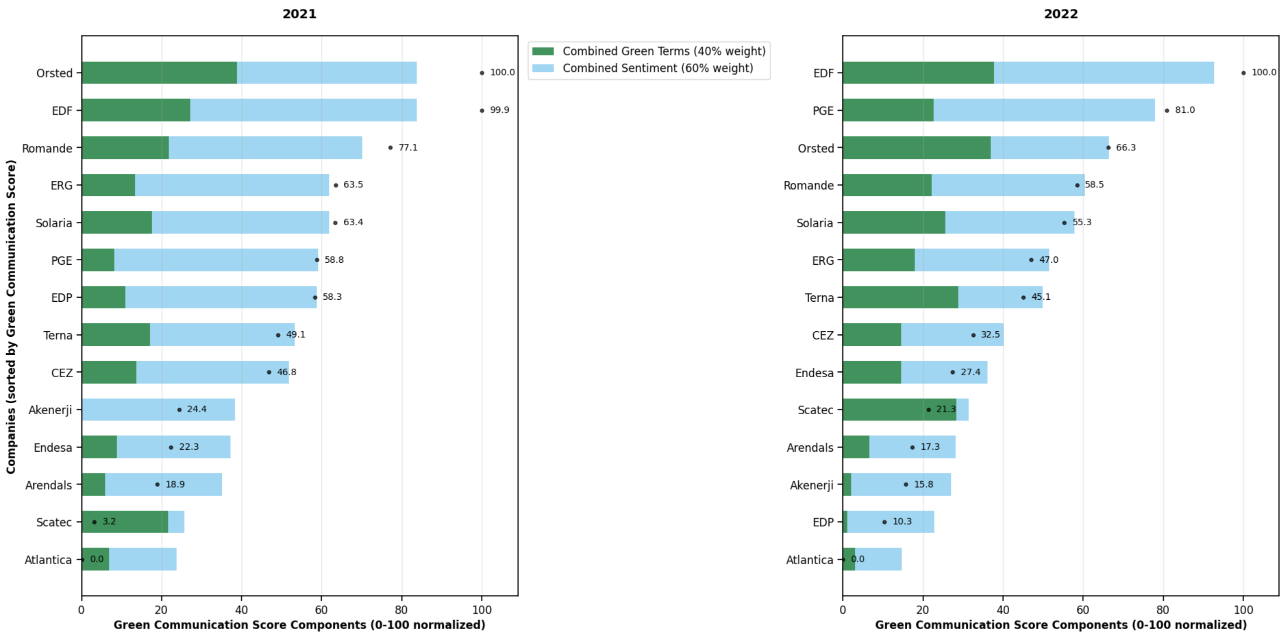

Figure 4.3: Green Communication Intensity

The green term component (green bars in *Figure 4.3*) shows substantial variation, with companies like Ørsted and EDF achieving high term frequency while Atlantica using minimal environmental vocabulary (Appendix F). The sentiment component (blue bars) plays the largest role in the final score, with a weight of 60 percent. Even the lowest sentiment scores remain above zero. This is due to the positive baseline produced by ClimateBERT's climate-specific sentiment analysis, which tends to classify environmental language as positive even when it refers to difficulties or limitations (see Chapter 3.4.3 for methodology details). The 40%-60% weighting is already reflected in the bar proportions shown in *Figure 4.3*.

The black dots in *Figure 4.3* represent the final normalised green communication scores (0-100 scale) that feed into the PCG. This additional normalisation step ensures comparability when combining communication intensity with the performance dimension, where 0 indicates the lowest communicator within this sample and 100 represents the highest. EDF maintains maximum intensity (99.9→100.0) through comprehensive environmental messaging, while Ørsted shows a large decline (100.0→66.3).

Table 4.1: Performance-Communication Change Alignment Analysis (2021-2022)

Company	Performance Change	Communication Change	Change Pattern
Akenerji	+26.39	-8.68	Not aligned
Arendals	+7.20	-1.64	Not aligned
Atlantica	-35.12	0.00	Not aligned
CEZ	-9.35	-14.22	Aligned
EDF	+10.00	+0.10	Aligned but disproportionate
EDP	+10.91	-47.95	Not aligned
Endesa	-16.79	-16.44	Aligned
ERG	+63.00	+5.10	Aligned but disproportionate
Ørsted	-24.48	-33.74	Aligned
PGE	+31.13	+22.17	Aligned
Romande	+22.50	-18.66	Not aligned
Scatec	-7.07	+18.08	Not aligned
Solaria	+12.41	-8.09	Not aligned
Terna	-7.05	-4.01	Aligned

Table 4.1 shows the absolute year-over-year changes in both performance and communication scores across the sample. Only five companies ($\tilde{3}6\%$) show aligned patterns where both dimensions change in the same direction with similar proportion. Companies like ERG show large performance gains (+63.0) alongside smaller communication increases (+5.1), while others like EDP show performance improvements (+10.91) paired with substantial communication decreases (-47.95). Scatec presents the reverse pattern, with declining performance (-7.07) accompanied by increased communication (+18.08).

Beyond communication intensity, the GRAT examines four additional communication quality dimensions.

4.2.2 Communication Quality Dimensions

Figure 4.4 shows how communication quality patterns appeared consistently across companies in both reporting years. The heatmap displays normalized scores (0-100) across five communication dimensions, where darker green indicates higher problematic scores and darker red shows lower risk patterns. Companies are sorted per year by their overall greenwashing risk levels, with the highest-risk companies (Romande, PGE, CEZ, ERG in 2021 and PGE, Ørsted, EDF and CEZ in 2022) appearing at the top. The colour intensity shows that most companies score above 50 points on temporal orientation (indicating future-focused messaging), while substantiation weakness shows the widest variation across companies, ranging from near-zero for companies like Atlantica and Akenerji to maximum scores for CEZ and Ørsted. These are not percentage scores but composite measures built from multiple subcomponents detailed in Chapter 3.4.3. For substantiation weakness, high scores indicate a stronger use of aspirational language rather than quantified or evidence-based claims. Higher language vagueness scores correspond to more frequent use of vague terminology. High temporal orientation scores correspond with future-focused commitments lacking specific timelines. Reporting consistency scores indicate year-over-year similarity suggesting template-based communication. Complete NLP module descriptions and verification appear in Appendix D.

Correlation analysis suggests that communication dimensions reflect related but distinct strategic choices, not separate constructs, with a moderate average correlation (r = 0.497) pointing to coordinated yet differentiated strategies across dimensions. Complete correlation analysis and statistical significance testing can be found in Appendix G.

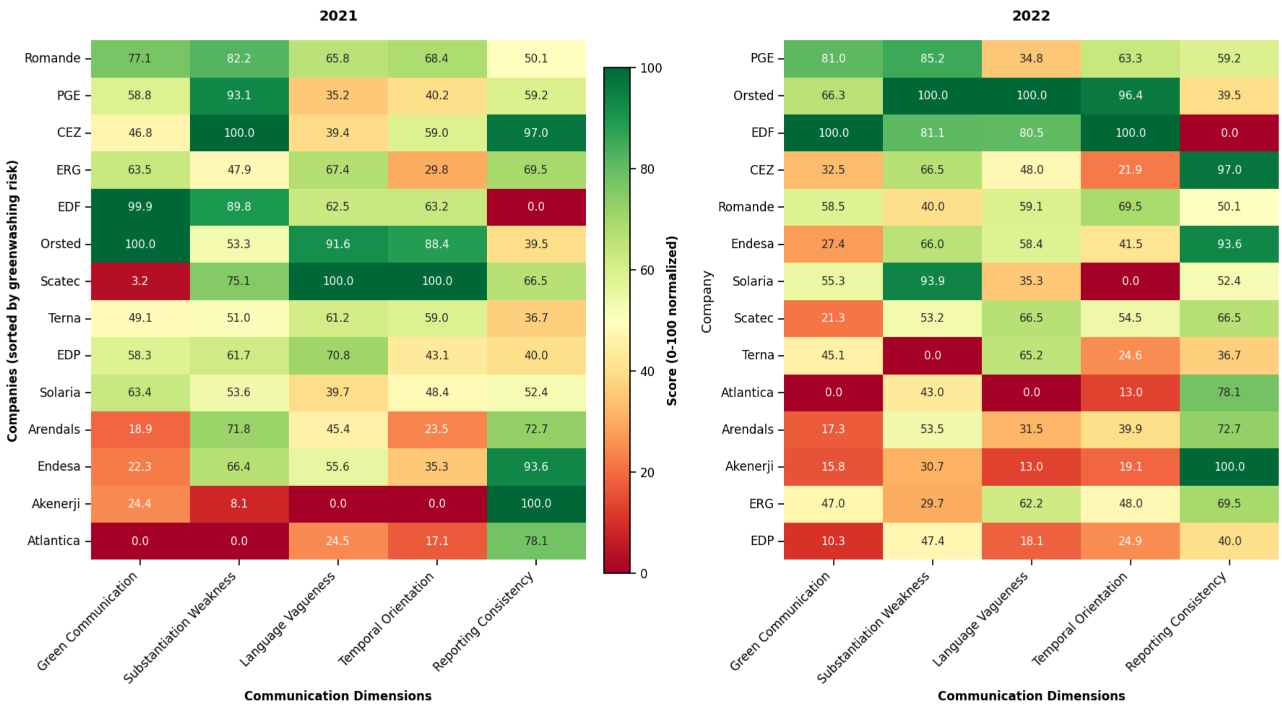

Figure 4.4: Multi-Dimensional Communication Profiles

Companies with high substantiation scores like CEZ (100.0 in 2021), Ørsted (100.0 in 2022), and PGE (93.1→85.2) show systematic use of aspirational instead of evidence-based claims, while those with lower scores like Atlantica (0.0→43.0) and Akenerji (8.1→30.7) show stronger evidence-based approaches despite their poor performance levels.

The reporting consistency results indicate that many companies rely on repeated, template-driven communication. High consistency scores indicate very similar communication across years. Akenerji (100.0), CEZ (97.0), and Endesa (93.6) used closely matching language between 2021 and 2022, as these scores reflect year-over-year textual similarity. Hence the exact same scores both years. Companies scoring lower on other communication quality dimensions often maintain high consistency scores, indicating systematic template usage rather than gradual changes in how companies report information. EDF's minimum consistency score (0.0) reflects substantial communication changes between years. This is due to the differences in their reporting structures between 2021 and 2022.

4.3 Performance-Communication Gap Analysis

Linking performance scores with green communication intensity applies Delmas and Burbano [48] core greenwashing concept by quantifying the PCG. *Figure 4.5* shows that performance-communication relationships form distinct strategic clusters rather than random distributions, using sample medians as reference lines to create quadrant-based risk assessment. The point sizes indicate the absolute gap magnitude, while red edges mark companies falling within the greenwashing risk zone.

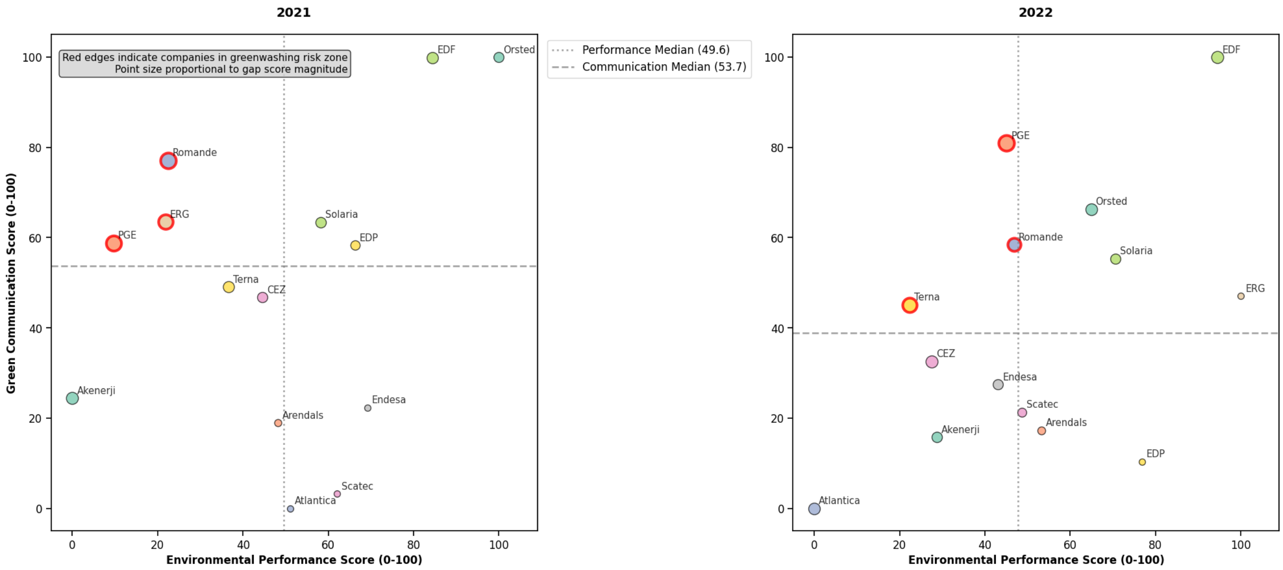

Figure 4.5: Performance-Communication Gap

The absolute gap between performance and communication scores receives a 1.5x amplification for companies positioned in the greenwashing risk zone (top-left quadrant), as detailed in Chapter 3.3.1. The amplification highlights the most critical disconnects between performance and communication, separating them from less severe cases. Table 4.2 presents the final PCG scores after amplification, where higher values indicate greater greenwashing risk patterns. This is the main (normalised) component for the greenwashing risk score detailed in 4.4. The gap scores range from perfect alignment (0.00 for EDP 2022 and Scatec 2021) to maximum risk (100.00 for PGE 2022 and Romande 2021). Eight companies show increasing gap scores between years, while the other six companies show decreasing gaps. The largest increases appear in Atlantica (+49.71), Terna (+33.00), and Endesa (+33.70), while the largest decreases occur in ERG (-74.86), Romande (-30.29), and Akenerji (-14.66). Companies maintaining consistently high gap scores include PGE (94.10→100.00) and CEZ (43.32→59.43), while companies achieving consistently low gaps include EDP (36.14→0.00) and ERG (86.09→11.23).

Table 4.2: Performance-Communication Gap Scores by company and year

Company	Perf.-Comm. Gap 2021	Perf.-Comm. Gap 2022
Akenerji	59.11	44.45
Arendals	21.01	25.40
Atlantica	5.54	55.25
CEZ	43.32	59.43
EDF	52.74	59.80
EDP	36.14	0.00
ERG	86.09	11.23
Endesa	8.54	42.24
Ørsted	41.76	56.38
PGE	94.10	100.00
Romande	100.00	69.71
Scatec	0.00	32.51
Solaria	45.41	42.55
Terna	50.61	83.61

PGE consistently occupies the highest-risk position, maintaining high green communication intensity despite low performance (as seen in *Figure 4.5*). PGE's consistent positioning points to a consistent PCG rather than a one-time deviation.

ERG shows a substantial shift in trajectory. Its movement from high-risk positioning to near-perfect alignment corresponds with performance gains (+63.0 points from *Figure 4.2*) that exceeded changes in communication intensity. These gap patterns form the foundation for the integrated risk assessment that combines dimensions.

4.4 Integrated Greenwashing Risk Assessment

This section addresses RQ3 by integrating environmental performance and communication characteristics to reveal greenwashing risk patterns. The analysis combines the five dimensions mentioned through ensemble methodology, producing risk scores that range from 16.5 to 81.3 across the sample.

The PCG identified in Section 4.3 is combined with four independent communication quality dimensions: substantiation weakness, language vagueness, temporal orientation, and reporting consistency. These five components are integrated using an ensemble method applied across 59,881 weight combinations (see Appendix H for statistics). *Figure 4.6* presents the final risk scores, based on the median values from all combinations. *Figure 4.7* breaks down how each component contributes to the final integrated scores.

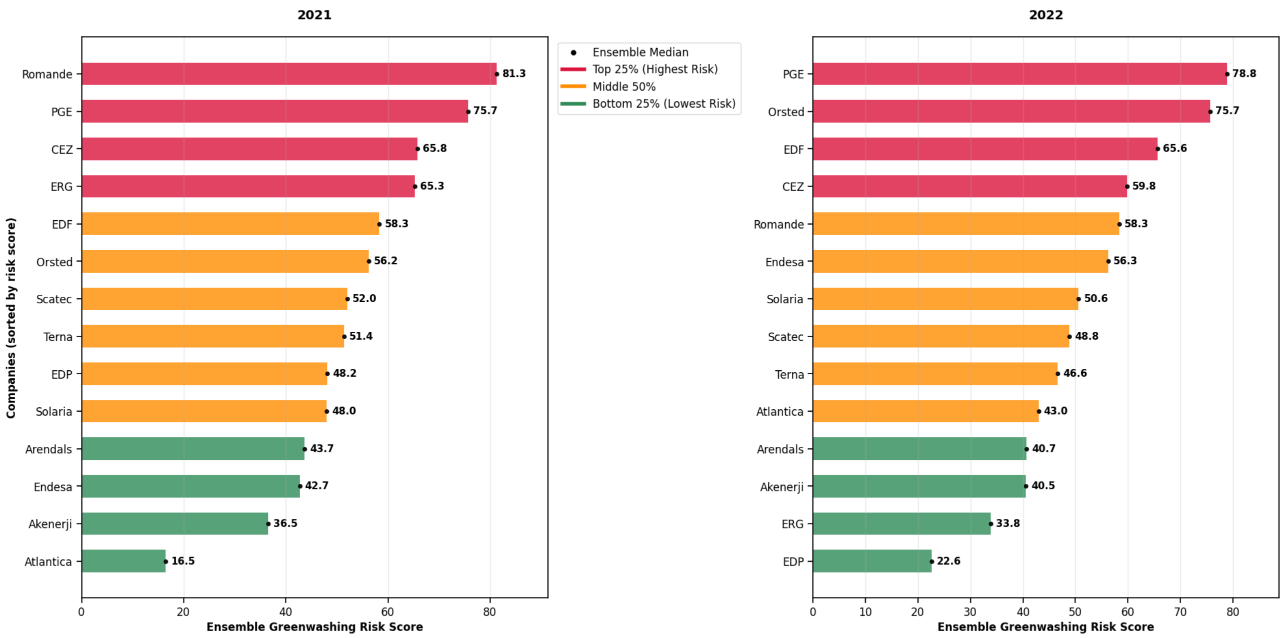

Figure 4.6: Final Greenwashing Risk Assessment

4.4.1 Risk Distribution and Methodological Robustness

Figure 4.6 shows the final greenwashing risk distribution across all companies and years. Using percentile-based groupings, companies are classified into top 25% (highest risk), middle 50%, and bottom 25% (lowest risk) categories. Notably, PGE and CEZ appear in the top 25% for both years, suggesting persistent greenwashing risk patterns. The ensemble methodology demonstrates strong stability, with uncertainty ranges from minimal variation (0.6 for Solaria 2021) to moderate uncertainty (5.7 for Scatec 2021), with most companies showing uncertainty ranges below 4 points (see Appendix

H). The final risk scores are unnormalised ensemble medians, preserving the combined effect of the five greenwashing dimensions. The scores range from 16.5 to 81.3, without companies reaching the limits of 0 or 100. This pattern suggests that the dimensions reflect different aspects of greenwashing behaviour rather than overlapping measures. Companies that score high on some dimensions often score lower on others, as shown in *Figure 4.7* and Appendix G.

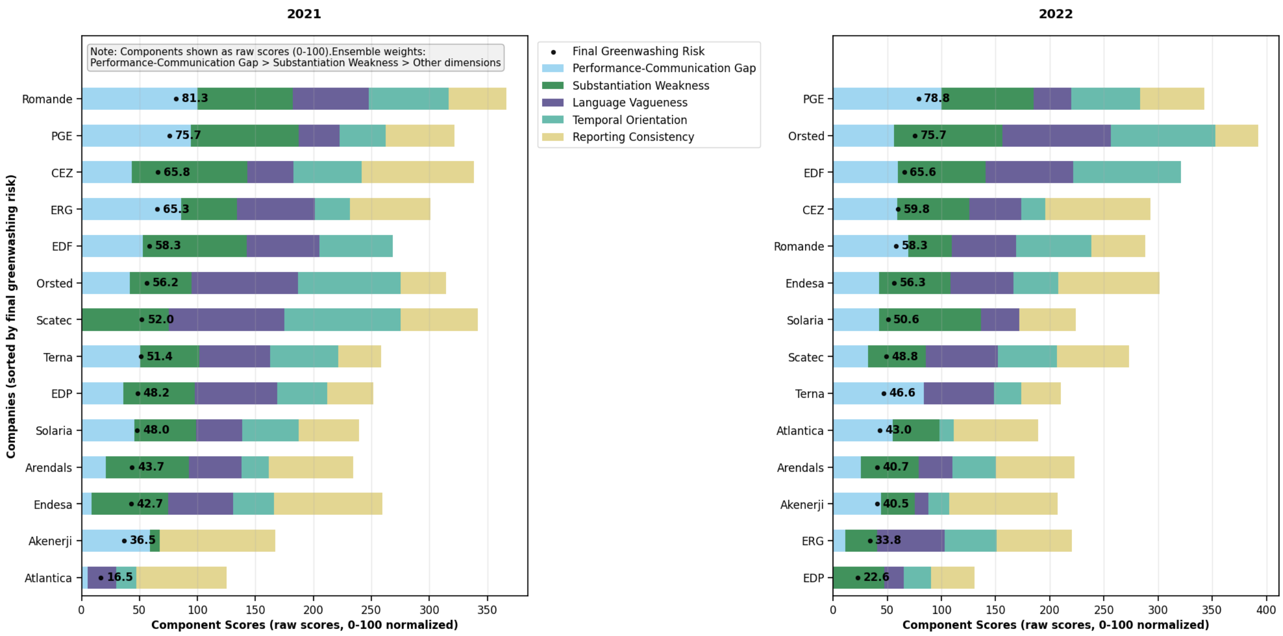

Figure 4.7: Greenwashing Risk Dimension Contributions

Figure 4.7 shows how individual dimension scores (visible as stacked coloured bars) combine to produce final risk scores (black dots) through theoretical weighting hierarchies. The stacked bars represent raw dimension scores before ensemble weighting, while the black dots show the final weighted scores (median scores across the variations). This is a similar plotting style as used for the performance components in *Figure 4.1*. Complete dimension scores and subcomponent variables are detailed in Appendix F. The PCG (light blue sections) carries the most weight in the final score, followed by substantiation weakness. The remaining communication quality components have smaller effects on the final score. The effect of weighting priorities becomes clear when comparing EDF 2021 (58.3) with Ørsted (56.2) and Scatec (52.0). Although Ørsted and Scatec display larger stacked bars, their overall scores are lower.

PGE's consistently high scores (75.7→78.8) come from large PCG contributions visible in *Figure 4.7*, combined with poor communication quality across multiple dimensions. The decrease in ERG's score from 65.3 to 33.8 is largely due to a reduced PCG.

5 Application of the GRAT

The Greenwashing Risk Assessment Tool (GRAT) provides systematic greenwashing detection for electric utility companies. The tool is specifically designed for this sector, but the user can adapt it for other industries by changing metrics and terminology. Regulators, investors, and researchers can apply GRAT directly to electric utilities or modify it for different sectors. Corporate Sustainability Reporting Directive (CSRD) implementation has standardised reporting recently, which makes application easier than during the pre-CSRD period analysed in the temporal scope. This chapter offers application guidance for future research without needing programming expertise. All code modules are available on GitHub (see Appendix K) and work independently once supplied with the right data inputs.

5.1 Understanding GRAT Architecture

GRAT combines five dimensions to detect greenwashing risk through integrated analysis rather than isolated metrics. This approach reveals patterns that single indicators miss.

Core Structure

- **Performance-Communication Gap (PCG):** Measures the fundamental disconnect between actual environmental performance and how companies communicate about this performance.
- **Four additional Communication Dimensions:** Analyse different aspects of how companies write reports. These dimensions work separately from performance and regulators consider them problematic.

Why This Works
Companies engaging in greenwashing typically show specific patterns, they:

- have poor environmental performance combined with positive environmental messaging.
- make vague promises without concrete evidence.
- use repetitive communication across years.

GRAT systematically detects these patterns by comparing what companies do (performance) with how they talk about it (communication).

The Integration Logic
GRAT uses ensemble methodology rather than relying entirely on fixed weights, which can overlook important patterns. Some fixed weights are still applied in specific calculations for the greenwashing sub-component; this is discussed in more detail later. This ensemble methodology tests all possible weight combinations within sound theoretical ranges. It then provides results showing both the most likely score (median) and the uncertainty range (standard deviation of all combinations). This approach prevents arbitrary decisions from affecting results. It also acknowledges that user priorities vary, offering flexibility to modify weights based on specific needs or preferences.

5.2 Data Requirements and Assessment Process

GRAT relies on systematic data collection from multiple sources to support integrated analysis of performance and communication. The GRAT follows a structured three-step process that turns raw inputs into usable greenwashing risk scores.

5.2.1 Essential Inputs

Performance assessment requires emission data from both Carbon Disclosure Project (CDP) (or equivalent self-reported sources) and Refinitiv Eikon (RE) (or other third party verified sources). Scope 1 and 2 emissions are minimum requirements. Third-party verification sources like RE Data provide the main analysis foundation. These sources need Scope 1 and 2 emissions calculated similarly to self-reported data. Scope 3 emissions play a meaningful role when measured consistently across companies. In this thesis, they were excluded due to lack of standardisation within the sample. Renewable energy amounts are particularly important for electric utilities, but due to missing data for some companies in the sample this was used as an additional score instead of a core component (discussed later). Emission and renewable energy data must use consistent intensity metrics across all companies and sources. Denominators might include revenue, MWh produced, or other relevant measures. This is necessary to standardise the data, allowing for a meaningful comparison. Target documentation should specify both target year and emission reduction percentage. Percentage reductions work best as they feed directly into existing calculations. RE Data includes this information, alternative sources work if they contain these required elements.

Temporal analysis for emission and renewable energy intensity improvement requires at least two consecutive years. Additional years improve reliability. All companies in the analysis must have an equal number of data points. Mixing companies with different observation periods undermines comparative analysis. For incomplete datasets, exclude companies that are missing essential performance metrics rather than estimating values. Replacing core emission data with industry averages produces unreliable results. However, for secondary elements like targets or renewable energy data (currently a bonus component), previous year values can substitute for current year gaps with proper documentation.

Communication analysis requires English-language sustainability reports in digital format for text processing. Non-English reports need professional translation, although this reduces reliability. Reporting periods must align across years to enable consistency analysis.

Table 5.1: Data Requirements Summary

Data Type	Requirements	Sources	Critical Notes
Performance	• Scope 1 and 2 emissions (Scope 3 preferred) • Renewable energy data • Intensity metrics (per revenue/MWh) • Targets with year and reduction percentage • Minimum 2 consecutive years	• CDP or equivalent self-reported sources • Refinitiv Eikon Data or other third-party verification	Exclude companies with missing core data. Scope 2 and Scope 3 calculation methods must be consistent across all companies
Communication	• English sustainability reports • Digital format • Consistent reporting periods	• Company websites • Regulatory databases	Professional translation required for non-English reports

5.2.2 Three-Step Assessment Process

The assessment process transforms raw data inputs into final greenwashing risk scores through systematic analysis. Users provide the required inputs while code modules handle all calculations automatically. The steps are performance scoring, communication analysis, and integration.

Step 1: Performance Scoring (0-100 scale)
Before beginning analysis, users must verify several data elements. Check emission data consistency across companies. Verify calculation method alignment (location-based vs. market-based). Validate intensity metric denominators. Confirm target documentation completeness.

The performance part of the GRAT then calculates four performance components. Emission intensity relative to sector benchmarks. Goal achievement rate compares actual progress vs. stated targets. Target ambition assesses reduction percentages against science-based targets. Transparency score evaluates differences between self-reported and third-party verified data.

The constrained ensemble methodology processes these components across 2,285 valid weight combinations. These weight combinations represent different ways to allocate importance among the four performance components while respecting theoretical priorities (emission intensity receives highest weight, followed by goal achievement, target ambition, and transparency). However, these combinations are based on chosen priorities drawn from the literature. Users can adjust the weight restrictions if needed, which leads to a different number of possible combinations. This generates performance scores with uncertainty statistics including mean, median, and Interquartile Range (IQR). Recommended is to use the median scores for further analysis in the next steps, as median values are less sensitive to outliers and provide more stable results across the range of weight combinations. A renewable energy bonus (up to +10 points) applies for strong renewable performance. A target change penalty (up to -10 points) applies for weakened commitments. All calculations can be seen in Chapter 3.3.1. For the actual code see the GitHub repository (Appendix K).

Step 2: Communication Analysis (Five Dimensions)
Text preparation begins with extracting text from PDF sustainability reports using SpacyLayout, again this is an existing code module that can be used directly. This creates .txt documents that serve as input for communication analysis modules. Users should verify if additional cleaning is required for their specific document formats. Sustainability reports are analysed across five dimensions (specified in Table 5.2) using rule-based Natural Language Processing (NLP) modules. Appendix D provides detailed technical specifications for each module. Table 5.2 summarises what each dimension measures and its associated red flags. Note that Green Communication scores combine with performance scores in Step 3 to create the Performance-Communication Gap (PCG), which becomes the primary greenwashing risk dimension.

Users simply provide text inputs, while the code takes care of normalisation and score integration on its own. However, users can modify component weights in the next step based on their preferences. Keep in mind that drawing reliable conclusions requires multiple rounds of expert consultation for weight validation.

Table 5.2: Communication Dimensions and Indicators

Dimension	What It Measures	Red Flags
Green Communication	Environmental term density and vocabulary diversity	Excessive environmental messaging compared to report length
Substantiation	Ratio of evidence-based and quantified claims to aspirational statements	High aspirational content with minimal evidence or quantification
Vagueness	Presence of hedging and uncertain language	Frequent conditional language: "may," "could," "aims to"
Temporal Focus	Balance between future commitments and present actions	Predominantly future-oriented with limited current achievements
Consistency	Year-over-year textual similarity	Above 95% similarity could indicate potential template usage

Step 3: Integration
Integration begins by calculating the Performance-Communication Gap (PCG) as the absolute difference between performance and communication scores. This transforms the Green Communication dimension into the PCG dimension by combining it with performance data. Companies in the greenwashing zone (poor performance combined with high green communication) receive a 1.5x amplification factor to highlight critical disconnects.

Final integration combines all five dimensions using ensemble methodology with 59,881 valid weight combinations. This maintains theoretical constraints where PCG receives highest weight and substantiation weakness receives second highest weight. The five integrated dimensions are: PCG (derived from performance and Green Communication), Substantiation, Vagueness, Temporal Focus, and Consistency. Final risk scores combine all dimensions through ensemble weighting, based on the priorities. PCG receives highest priority, followed by substantiation weakness. Remaining dimensions get weighted proportionally. Output consists of a risk score (0-100) with associated uncertainty range. The complete methodology exists in ready-to-use code modules, requiring only data inputs and basic execution capability.

For electric utility applications with complete renewable energy data across all companies, this component should transition from a bonus to a core metric. The renewable score should be positioned after goal achievement rate in the priority hierarchy. Standard implementation analyses two reporting years, requiring three years of performance data and two sustainability reports. Single-year analysis remains possible with two years of performance data, though two reports are still necessary for consistency detection through year-over-year comparison.

5.3 Interpreting Results and Validation

GRAT offers multiple validation methods to support reliable and accurate results. This becomes especially important when the tool is applied to new contexts or when default weight settings are adjusted. The tool's effectiveness improves with larger sample sizes. With more data, normalisation becomes more reliable, leading to stronger benchmarks and clearer relative risk scores.

Internal validation exists in the code and should be applied, especially when making weight modifications. The sensitivity analysis measures how component weight changes affect final scores and company rankings. This reveals whether GRAT produces stable results or if minor methodological adjustments lead to large shifts in outcomes. Robustness testing trough sensitivity analysis is essential for maintaining confidence in the results.

External validation builds on internal testing by comparing results against real-world greenwashing evidence. This involves testing against documented greenwashing cases within the sample and comparing rankings with expert assessments. The research method used for finding existing greenwashing cases can be applied to new contexts. The Mann-Whitney U (MWU) test can be implemented by changing the two comparison groups with the found evidence for the new sample. Statistical validation through MWU testing becomes viable with samples exceeding 30 companies. This makes this validation step increasingly powerful with larger sample sizes where statistical significance can be achieved. The current 14-company sample provides indicative rather than conclusive validation.

Ongoing validation keeps the tool reliable for repeated applications. This includes tracking prediction accuracy over time, especially relevant considering the changing CSRD-landscape. Update word lists based on changing corporate language. To keep the tool effective over time, it is important to perform regular quality checks and compare results with alternative methods as communication patterns evolve.

5.4 Sector Adaptation Guide

GRAT was built for electric utilities, but its architecture is flexible enough to apply in other sectors. The core process of measuring performance, analysing communication, and integrating results remains intact. What differs are the metrics and terminology, which must reflect sector-specific factors. Different sectors need adjusted performance metrics while keeping intensity-based comparability. Manufacturing sectors measure emissions per unit produced and need terminology related to production processes. Financial institutions focus on financed emissions or portfolio carbon intensity, with extra attention to disclosure practices. Retail businesses track supply chain emissions per revenue, which accounts for Scope 3 prominence. Transportation companies use fleet efficiency metrics and include technology transition terminology.

Table 5.3: Sector Adaptation Examples

Sector	Performance Metrics	Key Adjustments
Manufacturing	Emissions per unit produced	Terminology for production processes
Finance	Financed emissions/portfolio intensity	Disclosure quality emphasis
Retail	Supply chain emissions per revenue	Scope 3 prominence adjustment
Transport	Fleet efficiency metrics	Technology transition language

Calibration Requirements
Successful adaptation requires several elements. A minimum of 20 to 30 companies establishes sector baselines. The more companies added, the better the tool's results, as mentioned before. It uses sector averages for emission and renewable scores and normalises the greenwashing dimension scores based on sample size. More companies mean better risk detection.

To enable validation testing for different sectors or samples, documented greenwashing cases specific to each sector or sample must be identified. Expert consultation or sector knowledge guides weight selection. Sector-specific terminology lists capture relevant language patterns and green terms. The five-dimension structure remains intact while ensemble priorities adjust based on sectoral requirements. Adapting word lists requires sector-specific modifications to capture relevant terminology patterns.

Here are some examples:

- For manufacturing contexts, environmental language might include terms like "eco-efficient," "lean manufacturing," and "circular economy." Technology-related terms could include "automation" and "process optimization."
- Financial sector adaptations would incorporate language around "sustainable finance," "ESG integration," and "climate risk." Evidence-related terms include "portfolio screening" and "impact measurement."
- Retail sector modifications might focus on "sustainable sourcing," "supply chain transparency," and "packaging reduction." Particular emphasis on Scope 3 terminology like "supplier engagement" and "value chain emissions" is important.

The examples suggest how the tool could work, but they are not strict instructions. Each sector should adjust the approach to fit its own situation. The adaptation process involves five iterative steps:

- Assess sector materiality to identify the most important environmental issues.
- Adjust the communication weights to focus on the dimensions that matter most for greenwashing in a specific sector.
- Modify performance metrics with appropriate intensity denominators and benchmarks.
- Test the adapted tool using documented sector-specific greenwashing cases.
- Keep refining the tool based on what validation shows and what experts suggest.

5.5 Practical Application Summary

Quick Start Requirements

- Minimum two consecutive years of data
- Three data sources: third-party verification, self-reported metrics, and sustainability reports.
- English or professionally translated reports.
- Basic data analysis and coding capability for execution.

Output Specifications

- Risk score ranging 0-100.
- Company rankings within the analysed sample.
- Dimension-specific insights into risk drivers.
- Uncertainty ranges from ensemble methodology and sensitivity testing.
- Confidence levels indicating result reliability.
- Comparative analysis valid only within small samples.

- Multi-year applications enable temporal pattern detection.

What GRAT Does Not Cover

- Non-emission environmental issues (biodiversity, waste, water).
- Social and governance aspects of greenwashing.
- Absolute judgments about greenwashing for individual companies.

6 Discussion

The findings from Chapter 4 and the validation appendices (I and J) need interpretation within existing theoretical frameworks. This chapter examines how the results extend current greenwashing theory while addressing three literature gaps identified in the introduction: the PCG, the multi-dimensional communication gap, and the sector-specific gap.

This research addresses the PCG by developing systematic integration of quantitative environmental performance measurement with multi-dimensional communication analysis. This reveals that performance-communication relationships form distinct clusters rather than random distributions among the 14 companies analysed. The multi-dimensional communication gap gets addressed through communication analysis that captures layered strategies across substantiation weakness, language vagueness, temporal orientation, and reporting consistency. This establishes that communication quality profiles operate independently of performance levels. The sector-specific gap receives partial attention through the first targeted greenwashing detection tool specifically applied to electric utilities. However, the findings from these 14 companies cannot be generalised to the broader European electric utility sector because of sample size limitations, leaving this gap only partially addressed.

The discussion is split into two main areas with different purposes. Section 6.1 examines the GRAT as a developed tool, focusing on its design process, limitations, user requirements, and practical applications for stakeholders and researchers. This section focuses on the tool itself, explaining how it can be adapted to other sectors and how future users can validate it. Section 6.2 interprets the specific patterns observed among the 14 companies analysed, connecting these patterns to existing literature and exploring their implications for understanding greenwashing mechanisms. The exploratory nature of the conclusions, due to the limited sample, is acknowledged throughout this chapter.

6.1 GRAT Development and Application

This section reflects on the GRAT as the primary contribution of this research. It examines how the tool was developed, its limitations, and guidance for future applications.

6.1.1 Tool Development Process and Design Choices

The performance component of the GRAT addresses the credibility paradox by using multiple data sources. This part of the tool uses both CDP self-reported data and RE third-party verification. The industry benchmarks come from 27 companies within the RE dataset to provide broader understanding of the sector beyond the main data sample. The threshold-based scoring uses scoring functions suited for specific measurement needs to reduce how much outliers can influence performance assessments. For instance, companies performing 50% better than industry average receive full points, while those performing twice as poorly get zero points.

Currency standardisation uses historical exchange rates so companies can be compared fairly across different reporting currencies. The renewable energy bonus (up to +10 points) and suspicious target change penalties (up to -10 points) address sector-specific transformation patterns and target manipulation risks found in the literature. These design choices reflect the unique characteristics of electric utilities during energy transition periods.

For communication analysis, the multi-dimensional NLP analysis component of the tool builds on corporate environmental communication theories from Chapter 2.3. It captures communication strategies that single-dimensional approaches miss through five different theoretical foundations. These include selective disclosure theory for green communication intensity, Clarkson et al. [43] hard versus soft disclosure distinction for substantiation weakness, regulatory guidance for language vagueness, temporal manipulation patterns for temporal orientation, and symbolic compliance theory for reporting consistency. This theoretical grounding was used to create the five dimensions in this analysis.

To ensure transparency and sector alignment, the rule-based method relies on SpaCy rather than black-box machine learning approaches. ClimateBERT was chosen for sentiment analysis because general models frequently misinterpret sustainability content, leading to unreliable results. Within each dimension, component weighting follows structured logic that corresponds to the strength found in existing literature. The human-AI collaboration methodology for creating NLP code modules combines automation efficiency with domain expertise. Claude Sonnet was used during development while maintaining oversight for pattern verification and rule refinement. Despite being automated, the processes were built on and informed by established theoretical concepts from chapter 2.

6.1.2 GRAT Limitations and Constraints

Several technical limitations affect how accurately the GRAT captures environmental performance and communication patterns. The emission intensity and renewable energy benchmarks depend on RE's available data rather than true sector-wide benchmarks, which could skew the performance outcomes. Scoring thresholds come from literature and logic but also partially from the sample's distribution to avoid situations where all companies get zero or perfect scores.

English-only analysis misses communication patterns that might appear in companies' native languages. Moreover, focusing only on sustainability reports ignores other communication sources like websites, social media, and press releases that could reveal different messaging strategies, particularly those directed to consumers rather than regulators.

Several NLP issues affect accuracy. The SpacyLayout parser sometimes fails to correctly identify column structures within reports, causing sentences to mix and affecting sentence-level pattern detection. Time constraints and limited linguistic expertise may have led to NLP modules that do not fully capture the intended dimensions. Examples include missing double negation checking for green terms and incorrect negation filtering. Additionally, the pre-CSRD reporting period meant that report structures and lengths varied widely across companies (as seen in Appendix B).

Goal achievement analysis gets restricted to current versus previous year comparisons because of limited historical data availability. Companies already generating substantial clean energy face particular challenges as fewer opportunities exist for large emission reductions when baseline performance is already strong. This creates an unintended penalty for environmental leaders. The literature emphasises renewable energy's importance in evaluating utility performance, but missing data for some companies required treating this component as bonus rather than core assessment.

The GRAT identifies patterns consistent with misleading environmental communication but does not prove that companies intentionally deceive stakeholders. Although validation results show that the GRAT can distinguish between companies with documented accusations and those with clean records, the sample size and other limitations must be considered when interpreting results.

It is important to note that companies receiving high greenwashing risk scores should not be assumed to be engaging in greenwashing. This research faced time and data limitations that prevent firm conclusions about any specific company's environmental behaviour. GRAT is intended to highlight patterns that may require further investigation, not to serve as proof of misconduct. These limitations make individual accusations inappropriate, as the tool is built for systematic risk screening rather than definitive judgment.

6.1.3 Weight Uncertainty and Ensemble Methodology

The weight uncertainty problem in existing greenwashing frameworks gets addressed through constrained ensemble methodology that examines all valid weight combinations within theoretical constraints. This approach provides transparency about methodological uncertainty while maintaining theoretical grounding. Some sub-components still rely on fixed weights, which are informed by available evidence but not explicitly defined in the literature. The methodology allows for flexibility: rapid sector assessments can use manual weight adjustments based on context-specific priorities, while long-term applications require expert consultation across multiple rounds to define stable and validated weight settings.

The constrained ensemble approaches systematically explores weight combinations within theoretically justified boundaries, testing 2,285 valid combinations for the performance score and 59,881 combinations for the greenwashing risk scores. This methodology represents a methodological advance for composite indicator development, providing a solution to the weight selection problem identified by Nemes et al. [173]. Academic researchers receive an assessment tool showing how performance and communication analysis can work together systematically while handling methodological uncertainty. The constrained ensemble approach works as a template for developing composite indicators when theoretical guidance exists, but optimal weights remain unknown.

Future applications should validate sector-specific configurations using the methods detailed in Appendices I and J, especially as larger datasets allow for more robust validation procedures.

6.1.4 Validation Approach and Statistical Limitations

The GRAT's validation combines internal consistency testing with external validation against documented cases. Internal validation examines whether the tool produces consistent results across different weight combinations and component specifications. External validation tests discriminant validity through documented greenwashing case analysis, with the tool successfully separating companies with credible accusations from those with clean records among the 14 companies analysed.

However, significant validation limitations must be acknowledged. The MWU test applied for validation lacks statistical power because of the small sample size with only 14 observations. The current validation results are indicative rather than conclusive, as more observations would provide substantially better statistical foundation. The patterns observed between 2021 and 2022 could represent random phenomena rather than systematic differences.

Future applications with larger datasets would enable more robust validation through meaningful statistical power. The MWU test becomes viable with sufficient sample sizes in both comparison groups, including companies with documented greenwashing cases and those with clean records.

6.1.5 Practical Implementation and User Guidance

The GRAT was initially designed for the electric utility sector, but it can be converted to serve multiple user types with different requirements and applications. Detailed guidance on data requirements, assessment processes, sector adaptation, and practical implementation is discussed in Chapter 5. The tool serves as both a screening mechanism for identifying companies requiring detailed investigation and a methodological template for developing composite indicators when theoretical guidance exists, but optimal weights remain unknown.

Rule-based NLP methodology provides transparency and interpretability that machine learning approaches often miss in regulatory contexts. The tool produces normalised risk scores across multiple dimensions, with companies scoring relative to their peers within the sample. However, it identifies companies requiring investigation rather than providing conclusive evidence of greenwashing.

6.2 Results Interpretation and Generalisability

This section interprets the specific findings from the 14 European electric utility companies analysed, discussing patterns observed and their broader implications within the acknowledged limitations.

6.2.1 Performance-Communication Decoupling Patterns Among the 14 Companies

The analysis of these 14 companies reveals systematic decoupling between environmental performance and communication strategies that aligns with existing theoretical frameworks. Among this sample, communication intensity operates as an independent capability, separate from environmental performance outcomes. This pattern supports theoretical predictions of systematic decoupling between symbolic and substantive behaviour, where organisations separate formal claims from technical operations to maintain legitimacy [155].

Figure 4.3 (showing green communication intensity through term frequency and sentiment analysis) demonstrates the independent nature of communication strategies observed among these companies. EDF maintains maximum intensity (99.9→100.0) despite modest performance improvements, while ERG shows declining communication intensity (63.5→47.0) alongside substantial performance gains (+63.0 points). This suggests that communication strategies often work independently from actual environmental outcomes through systematic decoupling mechanisms [29].

The large increase of ERG's performance requires some additional attention. ERG's transformation reflects a documented strategic shift toward becoming an operator purely focussed on renewables such as wind and solar. According to their Annual Report [60], ERG completed the sale of its hydro assets and added over 500 Megawatt (MW) of new renewable capacity within the same year. Similar developments are mentioned in their 2021-2025 Industrial Plan [61], which projected 90% of €2.1 billion in investments toward renewables and reported that 400 MW were already under construction at the time the plan was launched in 2021. A combination of phasing out assets and grouping renewable energy under a single portfolio likely explains the sharp one-year performance improvement. Short-term structural portfolio changes like this create amplified performance increases because improvements in emission intensity performance carries over into other components of the GRAT. For example, the 2022 emission reductions compared to the previous higher baseline intensity (from 2021) drive goal achievement rates, while forward-looking renewable investments support maximum target ambition scores.

The credibility paradox, mentioned throughout the report, shows up clearly in cases like PGE. Their +31.1-performance improvement achieved perfect goal achievement despite maintaining relatively high emissions. According to their Integrated Report [189], revenue increased significantly as a result of higher energy prices, while absolute emissions decreased slightly. Although this likely explains the decrease in intensity, the actual change in intensity is disproportionately large (see Appendix C1) and is not well supported by the structural or methodological changes mentioned in the report. When emission intensity improves without obvious structural changes, it creates concerns about measurement credibility. This reflects the 'credibility paradox', where companies control both disclosure and verification [29].

Table 4.1 shows that 9 out of 14 companies (around 64 percent) show decoupled patterns, where changes in performance and communication move in opposite or disproportionate directions. Only 5 companies (around 36%) show aligned patterns where both dimensions change in the same direction with proportional magnitude. This decoupling happens in different ways: ERG shows large performance gains (+63.0 points) alongside much smaller communication increases (+5.1 points), while EDP shows the opposite with performance improvements (+10.91 points) paired with substantial communication decreases (-47.95 points). Scatec shows reverse decoupling, where performance decreases (-7.07 points) while communication increases (+18.08 points).

This disconnect suggests that communication strategies operate independently of operational environmental achievements within this sample, aligning with selective disclosure theory predictions where companies control information flow regardless of actual performance [140, 29].

Some interesting examples illustrate the disconnected patterns. Ørsted's case shows how companies with solid environmental performance can still experience increasing greenwashing risk through communication choices. The company's greenwashing risk score (*Figure 4.6*) increased from 56.2 to 75.7 while its environmental performance remained high among the sample. This fits within signalling theory predictions [220] about communication decisions working independently of underlying capabilities. Similarly, EDF's greenwashing risk trajectory (58.3→65.6) occurs despite almost perfect environmental performance scores, suggesting that risk increases get shaped by communication choices rather than underlying performance issues within this sample.

On the other end, Atlantica provides an example of communication withdrawal during poor performance periods. The company's already minimal green communication decreased even further from 2021 to 2022, combined with declining performance visible in *Figure 4.2*. This minimisation approach is consistent with existing research showing that institutional pressure may cause organisations to separate symbolic communication from actual operations [29]. In some cases, they may withhold disclosure when credible claims are hard to support, choosing to report less when performance weakens rather than exaggerate achievements [149].

The integrated GRAT confirms that communication strategies work separately from environmental performance through these decoupling mechanisms among the 14 companies analysed. While these patterns align with established decoupling theory, generalisation beyond this specific sample requires a larger sample or additional research.

6.2.2 Multi-Dimensional Communication Quality Analysis

The five communication dimensions show distinct patterns across the 14 companies that extend beyond simple performance-communication relationships. *Figure 4.4* shows how companies have varying communication quality profiles, with some showing high scores on certain dimensions while scoring substantially lower on others. This suggests that companies use different communication strategies rather than following a single approach.

The correlation analysis (Appendix G) shows the strongest significant relationship between Language Vagueness and Temporal Orientation ($r = 0.786$, $p < 0.001$), indicating that companies within this sample using vague environmental language also tend to make future-focused commitments without specific timelines. This pattern points to coordinated communication strategies where unclear language is combined with distant promises. Green Communication Intensity shows a significant negative correlation with Reporting Consistency ($r = -0.676$, $p < 0.001$), indicating that companies with higher environmental messaging tend to have less consistent year-over-year reporting. This negative relationship may also reflect the pre-CSRD period when standardised reporting formats were not yet required, allowing companies with high communication intensity to vary their reporting approaches more substantially.

Companies use different communication strategies across dimensions. Scatec in 2021 shows this pattern well, achieving the highest scores among all companies for Language Vagueness (100.0) and Temporal Orientation (100.0), with above-average scores for Substantiation Weakness (75.1) and Reporting Consistency (66.5), yet maintaining very low Green Communication intensity (3.2, the second lowest in the sample). This suggests a strategy of minimal environmental messaging combined with vague, future-focused language when environmental topics are addressed.

The PCG shows weak and non-significant correlations with all four communication quality dimensions ($r = 0.013$ to 0.182, all $p > 0.05$), supporting the multi-dimensional GRAT design by suggesting that performance-communication mismatches and communication quality issues may represent separate greenwashing mechanisms. However, the small sample size limits definitive conclusions about these relationships.

Several companies consistently show lower greenwashing risk through their communication approaches across both years. Atlantica, Akenerji, and Arendals show relatively modest scores across most communication quality dimensions while also maintaining relatively low PCG scores. This suggests more conservative communication strategies that avoid the problematic patterns captured by the quality dimensions.

6.2.3 Generalisability Limitations

As discussed before, the findings from these 14 companies within the EU electric utility sector cannot be generalised to the broader European electric utility sector because of several limitations. With only 14 companies, the sample lacks sufficient statistical power to support broad generalisations about European electric utilities. The MWU test results acknowledge very low statistical power because of the small number observations. Even if patterns appear meaningful within this sample, they could represent random phenomena occurring between 2021 and 2022 rather than systematic sector-wide trends.

Sample selection was driven entirely by data availability rather than representativeness. Only companies with complete data across all three sources (CDP, RE, sustainability reports) were included, with CDP participation being voluntary and thus favouring companies willing to disclose environmental data. This creates systematic bias where companies with better data management practices may also exhibit different greenwashing behaviours compared to the broader sector. The implications of this systematic bias limit the representativeness of the findings.

The 2021-2022 analysis period represents troubling transitional years marked by COVID-19 impacts and pre-CSRD implementation conditions. COVID-19 created operational challenges that likely influenced both environmental performance and communication strategies in ways that may not reflect normal operating conditions. Additionally, the pre-CSRD period featured varied reporting structures across companies, which may not represent current standardised reporting practices following CSRD implementation. EDF is the most prominent example of this, transitioning from a brief impact report in 2021 to a large sustainability report in 2022, which resulted in unrealistically low reporting consistency scores.

The analysis focused on Scope 1 and 2 emissions while excluding Scope 3 emissions because of data availability and emission-calculation constraints. For electric utilities, Scope 3 emissions often represent significant portions of total environmental impact, particularly for companies involved in fossil fuel supply chains. This exclusion may affect performance assessments and limit the greenwashing risk evaluation.

Interpreting risk scores has real constraints given the sample limitations and time scope. Normalisation gets affected by sample size, meaning the small number of companies may skew how greenwashing dimensions get represented. This possibly resulted in rankings that misrepresent the utility sector as a whole. Even if all 14 companies perform reasonably well environmentally, the GRAT still identifies the 'worst' performers within this group, although they may not represent poor performance compared to the broader European sector.

6.2.4 Methodological Validation and Sector Implications

The internal validation results among these 14 companies provide exploratory evidence for the GRAT's methodological stability. The ensemble methodology shows potential for broader use, while also acknowledging the current limitations. The average standard deviation of the ensemble scores is 3.0 for the greenwashing risk score and 3.1 for the performance score, with scores ranging from 0 to 100 among these companies. This suggests that performance scoring and risk identification remains relatively stable across different weight combinations within this sample.

Sensitivity Analysis Results
Internal validation through sensitivity testing of the chosen fixed weights shows strong robustness within this sample. All 17 weight variation scenarios produced average rank shifts well below the 2.0 threshold for high robustness [201]. The scenarios with highest impact include green communication adjustments, which show the highest CV values (1.618-1.657) and maximum score shifts (1.72-1.73), indicating that green communication intensity weights most significantly influence score differences. The amplifier weight reduction (-10%) shows the highest rank shift impact (0.214) but remains well within acceptable bounds. Companies typically shift less than one position when communication subcomponent weights change by ±10% within this sample.

The company-level patterns show that four companies exhibit high score sensitivity (CV values above 15%), but their average rank shifts remain minimal, preserving the GRAT's ability to compare companies reliably within this sample. Companies with documented greenwashing histories show low to moderate sensitivity levels: PGE maintains the lowest sensitivity (CV = 2.1%), while CEZ and Ørsted fall into the moderate category.

Sector Analysis Implications
Despite generalisability limitations, the analysis provides insights into how systematic greenwashing detection could work within regulated utility environments. Companies within regulated environments may develop communication strategies that operate independently from operational environmental improvements. Electric utilities face unique pressures related to infrastructure investment timelines, regulatory approval processes, and public service obligations, potentially creating sector-specific communication patterns.

6.2.5 Theoretical Implications for Greenwashing Understanding

These results offer early insights into greenwashing patterns that can inform theory. Still, broader validation is necessary before making definitive claims. The findings challenge common assumptions about greenwashing by revealing that communication strategies work separately from environmental performance through decoupling mechanisms among this specific sample. Multi-dimensional communication analysis proves more effective at capturing misleading behaviour than single-dimensional approaches within these companies.

This represents a potential theoretical contribution, suggesting that greenwashing happens through multiple pathways beyond simple environmental overstatement: communication withdrawal strategies and temporal messaging patterns that focus on future commitments without near-term plans. Different companies within this sample handle institutional pressures and complexity in separate ways, as shown by the performance patterns. This observation could be relevant to organisational theory, especially in explaining how regulated sectors respond to legitimacy demands during operational transitions.

If confirmed in larger samples, the finding that green communication intensity functions as an independent capability could reshape how stakeholders interpret corporate environmental claims. The patterns observed among these 14 companies suggest that stakeholders may need more sophisticated frameworks for evaluating environmental claims that go beyond simple performance metrics. However, all theoretical implications remain exploratory given the sample constraints and transitional period context (COVID-19 and pre-CSRD implementation).

7 Conclusion

This research developed the GRAT specifically designed for electric utilities and tested it on 14 European electric utility companies. The GRAT represents the primary contribution of this thesis, addressing three fundamental gaps in existing literature through systematic integration of environmental performance measurement with multi-dimensional communication analysis. This concluding chapter addresses each research question directly, reflects on the tool's key features, and outlines clear directions for future use and research.

7.1 Answering the Sub-Research Questions

Each sub-research question addresses a different aspect of developing the GRAT. RQ1 establishes the environmental performance measurement foundation, RQ2 examines communication strategy patterns, and RQ3 integrates these domains to reveal greenwashing risk patterns.

7.1.1 RQ1: Performance Variations Among the 14 European Electric Utilities

Research Question 1: *What environmental performance variations exist among European electric utility companies based on systematic analysis of their available sustainability data?*

The 14 European electric utilities analysed show wide variation in environmental performance. Ensemble median scores range from 15.0 to 95.0 out of 100 points, with year-over-year changes spanning –35.1 to +63.0 points. The analysis reveals that performance variations follow recognisable patterns rather than random fluctuations. Companies achieve environmental performance through different pathways across emission intensity, goal achievement, target ambition, and transparency components, each contributing according to theoretical hierarchies established in literature. The most notable changes include ERG's transformation (+63.0 points), reflecting documented business model shift toward pure renewables operation, and Atlantica's decline (-35.1 points), demonstrating fluctuations in performance for infrastructure-focused companies as well. PGE's improvement (+31.1 points) achieved perfect goal achievement despite maintaining relatively high emission intensity, illustrating how revenue-based intensity calculations can create misleading progress indicators if based on performance from the previous year.

The four performance components show different distribution patterns. Emission intensity scores demonstrate the widest variation, from perfect scores (100.0 points) for companies like Ørsted and EDF to minimal scores for PGE and Akenerji (0.0-5.43 points). Goal achievement reveals an interesting credibility paradox, where companies control both disclosure and verification. Perfect scores (100.0 points) appear across very different company types: from clear transition leaders like EDF to coal-heavy companies such as PGE in 2022. This suggests that the way how companies set their targets may matter more than their actual emission reductions. Target ambition shows a shift toward stronger commitments across the sample, with fifteen (out of 28) company-year observations reaching maximum scores. However, this ambition comes at the cost of goal achievement scores as more than half of these companies have goal achievement scores below 35 (most are zero). Suggesting that companies' high target ambition often happens without matching implementation progresses. Companies receive relatively high transparency scores, reflecting how consistent disclosure is between CDP self-reported and RE third-party verified data.

While these findings provide insights into utility performance measurement approaches, generalisation beyond these specific companies requires larger, more representative samples.

7.1.2 RQ2: Communication Patterns Among the 14 European Electric Utilities

Research Question 2: *What multi-dimensional communication patterns characterise environmental disclosure in European electric utility companies?*

The communication analysis among these 14 companies shows that companies use different strategic approaches across five dimensions. These approaches work independently of performance levels. Companies show different communication quality profiles - some score high on certain dimensions while scoring much lower on others.

Key patterns include strong connections between Language Vagueness and Temporal Orientation ($r = 0.786$, $p < 0.001$) and negative connections between Green Communication Intensity and Reporting Consistency ($r = -0.676$, $p < 0.001$). Companies like Scatec show this variety well, achieving maximum scores for Language Vagueness (100.0) and Temporal Orientation (100.0) while keeping very low Green Communication intensity (3.2). This suggests minimal environmental messaging combined with vague language when environmental topics get addressed.

The communication dimensions show that companies use different strategic approaches across the five areas. Green Communication connects with problematic communication patterns - companies using more environmental messaging also tend to use more vague language, aspirational claims, and future-focused commitments. Green Communication shows a strong negative connection with Reporting Consistency ($r = -0.676$), indicating that higher environmental messaging goes with less consistent reporting. Companies fall into distinct patterns. Some like Atlantica, Akenerji, and Arendals consistently score lower across communication dimensions, using more conservative approaches. Others like CEZ and Ørsted score high across multiple problematic dimensions, combining different problematic communication strategies together.

7.1.3 RQ3: Greenwashing Patterns Among the 14 European Electric Utilities

Research Question 3: *What greenwashing risk patterns are revealed when environmental performance and communication characteristics are analysed together?*

The integration analysis of 14 companies within the sample reveals that 64% display decoupled patterns where communication and performance changes move in opposite or disproportionate directions, with only 36% showing aligned patterns where both dimensions change in the same direction with proportional magnitude. The PCG serves as the primary greenwashing risk indicator. It successfully identifies disconnects between environmental achievements and communication intensity.

The analysis reveals multiple greenwashing pathways beyond simple environmental overstatement. These include communication withdrawal strategies (as shown by Atlantica's case) and temporal messaging patterns that focus on future commitments without near-term implementation plans. Specific examples include ERG showing aligned but disproportionate changes (+63.0 performance gains, +5.1 communication increases). EDP demonstrates opposite-direction decoupling (+10.91 performance, -47.95 communication). Scatec presents reverse decoupling (-7.07 performance, +18.08 communication).

Companies achieve low greenwashing risk through different approaches: some minimise PCG contributions (EDP, ERG), while others use limited green communication intensity to avoid gap-related risk (Atlantica). The analysis demonstrates that communication quality problems amplify PCG effects, supporting multi-dimensional detection approaches. Importantly, the PCG shows weak and non-significant correlations with communication quality dimensions ($r = 0.013$ to 0.182), suggesting that performance-communication mismatches and communication quality issues represent distinct greenwashing mechanisms that can occur independently. These patterns align with theoretical expectations about decoupling mechanisms. However, the small sample size limits broader generalisation and keeps the findings exploratory.

7.2 Answering the Main Research Question

Main Research Question: *What greenwashing risk assessment tool can be developed for European electric utility companies based on environmental performance metrics and multi-dimensional communication analysis?*

This research successfully developed the GRAT. The tool systematically integrates environmental performance measurement with multi-dimensional communication analysis for greenwashing risk assessment. The GRAT operates through two integrated components. The first is a performance measurement system using quantitative environmental data. The second is a communication analysis system examining five distinct dimensions through rule-based NLP methods.

The GRAT addresses three fundamental literature gaps through specific methodological innovations:

- It bridges the PCG by systematically combining quantitative environmental data with qualitative text analysis through the PCG calculation.
- It addresses the multi-dimensional communication analysis gap by examining communication across substantiation weakness, language vagueness, temporal orientation, and reporting consistency using sector-specific rule-based methods.
- It provides targeted analysis for electric utilities while maintaining adaptability for other sectors through component modification.

The methodological framework uses transparency-focused design choices that enable regulatory and academic applications. Rule-based NLP approaches provide interpretability compared to machine learning methods, while the constrained ensemble approach handles weight uncertainty by systematically examining valid combinations within theoretical constraints rather than relying on arbitrary selections. The GRAT was validated through both internal consistency testing and external validation against documented greenwashing cases. However, the sample size constraints and other methodological limitations discussed in Section 6.1.4 affect these results. The external validation results should be considered indicative rather than conclusive.

For practical implementation, the tool serves multiple user types. These range from regulatory screening to academic research applications. It identifies companies requiring detailed investigation rather than providing conclusive greenwashing evidence. This makes it suitable for systematic analysis of large document sets while maintaining efficiency. Detailed implementation guidance appears in Chapter 5, allowing users to apply the GRAT to new datasets and adapt it for different sectors.

7.3 Future Research Directions

The limitations discussed in Section 6.1.2 highlight several constraints of the current GRAT implementation that future research can address, this is mentioned both here and in the following section.

More recent data and larger datasets would improve how accurately normalization works and should reveal clearer greenwashing patterns. Companies face growing pressure to disclose more as regulatory requirements and stakeholder demands increase. The potential for this approach keeps expanding. Recent regulatory developments make this scale expansion especially relevant. After CSRD implementation, research could benefit from standardized reporting structures. Technical challenges would decrease while providing more relevant regulatory insights. Comparative studies could examine the same companies before CSRD implementation (2021-2022) and after implementation (2025+). Whether stricter reporting requirements effectively reduce greenwashing behaviour could be assessed through this approach.

These larger and expanded datasets also enable sector-specific improvements and provide more robust statistical foundations for validation. For utilities specifically, renewable energy performance should receive greater weight based on literature support rather than being treated as bonus scoring because of data constraints. Including more companies in recent datasets would address this limitation while providing sector-specific insights. One of the noted limitations is the exclusive focus on sustainability reports. Researchers could address this by combining the GRAT with a web scraping tool that can extract large volumes of company communication from multiple sources. The NLP modules are already suited to handle this input without further modification.

Hybrid approaches that combine rule-based analysis, which is mostly used in this research, with machine learning could capture subtle linguistic strategies better while keeping results interpretable. Machine learning models could be trained on a large dataset of confirmed greenwashing cases across multiple years, using company communication as training input for the model. Once trained, these models could be applied to predict greenwashing risk in other companies. Communication patterns missed by English-only approaches would be revealed through multi-language analysis. Current pattern detection problems could be addressed through better NLP modules developed with greater linguistic expertise. Cross-sector comparative studies using adapted GRAT versions could identify sector-specific greenwashing patterns while testing the tool's broader applicability. How stakeholders respond to greenwashing detection needs more detailed examination, since this is still a literature gap. This would reveal in more detail how these findings translate into investment behaviour and policy responses.

7.4 Final Recommendations

The development of GRAT allows for several practical applications across various user groups. The following recommendations show how the tool can be used in policy, academic, and stakeholder contexts.

Policy and Regulatory Applications
Policy and regulatory development can benefit from several recommendations that come from this GRAT. European regulators should consider integrating multi-dimensional greenwashing detection into CSRD compliance monitoring. Large volumes of sustainability reports can be processed systematically through the GRAT, making it well-suited for regulatory oversight at larger scale. Instead of manual review, the GRAT could identify companies requiring detailed investigation. Limited resources could get directed more effectively this way.

The EU Green Claims Directive requires "robust, science-based evidence" to support environmental claims. The substantiation weakness dimension aligns directly with these requirements. Compliance with evidence-based disclosure requirements could be assessed through this component. Companies with high substantiation weakness scores rely systematically on aspirational rather than evidence-based claims. This indicates potential non-compliance with the evolving regulatory standards. Moreover, even when companies score low on substantiation weakness, this means their quantified claims can be directly compared to actual performance, making fact-checking easier.

Companies making distant promises without near-term implementation plans get revealed through the temporal orientation dimension. This pattern matches regulatory concerns about net-zero commitments that lack credible short-term action. Companies with high temporal orientation scores could be required to provide detailed targets and implementation milestones as part of regulatory approval processes. Documented greenwashing cases validate the GRAT, suggesting it could support risk assessment before enforcement actions. Companies with consistently high-risk scores and specific communication quality problems could attract closer oversight from regulators. Rule-based methods provide transparency that makes findings defensible in contexts that require explainable decision-making.

The GRAT provides a standardized approach for comparing greenwashing risk across European markets, serving international coordination purposes. As sustainability disclosure requirements expand globally, similar methodologies could help harmonize greenwashing detection. These approaches would need to account for both sector-specific and regional differences.

Academic and Methodological Applications
The GRAT shows how performance and communication analysis can work together systematically while handling methodological uncertainty. The constrained ensemble approach works as a template for developing composite indicators when theoretical guidance exists, but optimal weights remain unknown. This contributes to composite indicator methodology beyond greenwashing detection. Research institutions can adapt the tool for sector-specific studies while keeping methodological rigor. The rule-based NLP methodology offers transparency advantages over machine learning approaches in regulatory contexts. This makes it suitable for academic applications that need interpretable results.

Stakeholder and Investment Applications
Investment managers and ESG analysts can use the GRAT to screen environmental claims across portfolio companies. The tool shows specific communication quality issues that need further investigation. This helps allocate due diligence resources more efficiently and supports better investment decisions. Corporate sustainability managers can apply the tool to compare their communication practices with sector peers and find areas for improvement. The detailed insights across different dimensions help companies fix specific communication quality problems. This supports better environmental disclosure practices. The GRAT represents a significant step forward in systematic greenwashing detection while being transparent about methodological limitations. It is the first integrated tool designed specifically for combining performance and communication analysis. This gives stakeholders and regulators evidence-based methods for identifying potential environmental misrepresentation. It also contributes to theoretical understanding of greenwashing mechanisms.

Bibliography

[1] R. Adcock and D. Collier. "Measurement validity: A shared standard for qualitative and quantitative research". In: *American Political Science Review* 95.3 (2001), pp. 529–546. DOI: 10.1017/S0003055401003100.

[2] M. Adomßent and J. Godemann. "Sustainability communication: An emerging field in sustainability science". In: *Sustainability Science* 6 (2011), pp. 7–22. URL: https://link.springer.com/article/10.1007/s11625-010-0112-z.

[3] M. A. Ahmad, C. Eckert, and A. Teredesai. "Interpretable machine learning in healthcare". In: *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2018, pp. 559–560. DOI: 10.1145/3233547.3233667.

[4] G. A. Akerlof. "The market for "lemons": Quality uncertainty and the market mechanism". In: *The Quarterly Journal of Economics* 84.3 (1970), pp. 488–500. DOI: 10.2307/1879431.

[5] J. Amores-Salvadó, G. Martín-de Castro, and J. E. Navas-López. "CSR decoupling: When communication exceeds action". In: *Corporate Social Responsibility and Environmental Management* 30.1 (2023), pp. 45–59. URL: https://onlinelibrary.wiley.com/doi/10.1002/csr.2405.

[6] I. Antonopoulos et al. "Electricity balancing challenges for markets with high variable renewable generation". In: *Renewable and Sustainable Energy Reviews* 186 (2024), p. 113667. DOI: 10.1016/j.rser.2023.113667.

[7] D. I. Armstrong McKay et al. "Exceeding 1.5°C global warming could trigger multiple climate tipping points". In: *Science* 377.6611 (2022), eabn7950. DOI: 10.1126/science.abn7950.

[8] P. Arocena, A. G. Gómez-Plana, and S. Peña. "A model for the competitive benchmarking of energy costs". In: *Energy Economics* 131 (2024), p. 107385. DOI: 10.1016/j.eneco.2024.107385.

[9] B. Ayan, S. Abacıoğlu, and M. P. Basilio. "A Comprehensive Review of the Novel Weighting Methods for Multi-Criteria Decision-Making". In: *Information* 14.5 (2023), p. 285. DOI: 10.3390/info14050285.

[10] A. Bådagård, S. Piili, and P. Toivanen. *Don't get caught greenwashing: Lessons learned from a review of recent consumer ombudsman decisions on environmental claims*. 2024. URL: https://www.lexology.com/library/detail.aspx?g=5eba2fed-f87d-4cc6-bbe9-6821e56c6b15.

[11] M. Baldwin and J. Lammers. "Past-focused environmental comparisons promote proenvironmental outcomes for conservatives". In: *Proceedings of the National Academy of Sciences* 113.52 (2016), pp. 14953–14957. DOI: 10.1073/pnas.1610834113.

[12] A. Batmaz and G. Sisman-Aydin. "Türkiye's alignment with the Paris Agreement: a comparative policy analysis with Germany and Spain". In: *Sustainability* 17.9 (2025), p. 3899. DOI: 10.3390/su17093899.

[13] A. E. Beaton. "Overview of the scaling methodology used in the national assessment". In: *Journal of Educational Measurement* 29.2 (1992), pp. 112–122. DOI: 10.1111/j.1745-3984.1992.tb00372.x.

[14] T. E. Becker, M. M. Robertson, and R. J. Vandenberg. "Nonlinear transformations in organizational research: Possible problems and potential solutions". In: *Organizational Research Methods* (2019). DOI: 10.1177/1094428118775205.

[15] F. Bernini, M. Giuliani, and F. La Rosa. "Measuring greenwashing: A systematic methodological literature review". In: *Business Ethics, the Environment & Responsibility* (2023). URL: https://onlinelibrary.wiley.com/doi/full/10.1111/beer.12631.

[16] M. Bessec and J. Fouquau. “Environmental sentiment and stock prices: Evidence from the Wall Street Journal”. In: *Energy Economics* 92 (2020), p. 104960. DOI: 10.1016/j.eneco.2020.104960.

[17] E. Bevacqua, C.-F. Schleussner, and J. Zscheischler. “A year above 1.5 °C signals that Earth is most probably within the 20-year period that will reach the Paris Agreement limit”. In: *Nature Climate Change* 15 (2025), pp. 262–265. DOI: 10.1038/s41558-025-02246-9.

[18] L. Bindi. “Designing a semi-automated approach to find potential evidence of corporate greenwashing”. TU Delft Repository. Master’s thesis. TU Delft, 2024.

[19] J. Bingler, M. Kraus, and M. Leippold. “Cheap talk and cherry-picking: What climate target transparency reveals about greenwashing”. In: *Management Science* 69.4 (2023), pp. 2023–2048. DOI: 10.1287/mnsc.2022.4503.

[20] J. Bingler et al. “Leveraging language models to detect greenwashing”. In: *arXiv preprint arXiv:2311.01469* (2023). URL: https://arxiv.org/abs/2311.01469.

[21] A. Bjorn et al. “Determinants of corporate climate change mitigation targets in major United States companies”. In: *Climate Policy* 17.3 (2017), pp. 313–326. DOI: 10.5539/jsd.v10n1p71.

[22] A. Blackman. “Transparency for governance: The mechanisms and effectiveness of disclosure-based and education-based transparency policies”. In: *Global Environmental Change* 22.1 (2012), pp. 32–42. DOI: 10.1016/j.ecolecon.2011.03.006.

[23] BloombergNEF. *Global Investment in the Energy Transition Exceeded $2 Trillion for the First Time in 2024*. 2025. URL: https://about.bnef.com/insights/finance/global-investment-in-the-energy-transition-exceeded-2-trillion-for-the-first-time-in-2024-according-to-bloombergnef-report/.

[24] F. Boelders. “The relationship between firm characteristics and greenwashing in the European energy sector: An NLP approach”. Master’s thesis. Utrecht University, 2020. URL: https://studenttheses.uu.nl/handle/20.500.12932/37014.

[25] P. Bourke and D. Andrich. “Performance indicators: Just how do you weight them?” In: *Higher Education Research and Development* 13.2 (1994), pp. 169–181. URL: https://ro.uow.edu.au/articles/journal_contribution/Performance_indicators_Just_how_do_you_weight_them_/27718608?file=50470335.

[26] F. Bowen and J. A. Aragon-Correa. “Greenwashing in corporate environmentalism research and practice”. In: *Organization & Environment* 27.2 (2014), pp. 107–112. DOI: 10.1177/1086026614537078.

[27] A. Bradford. *The Brussels effect: How the European Union rules the world*. Oxford University Press, 2020. URL: https://scholarship.law.columbia.edu/books/232/.

[28] Brightest. *The EU Corporate Sustainability Reporting Directive (CSRD) - Timeline and Standards Overview*. 2024. URL: https://www.brightest.io/csrd-corporate-sustainability-reporting-directive.

[29] P. Bromley and W. W. Powell. “From smoke and mirrors to walking the talk: Decoupling in the contemporary world”. In: *Academy of Management Annals* 6.1 (2012), pp. 483–530. DOI: 10.5465/19416520.2012.684462.

[30] C. Bucur et al. “Sentiment analysis of global news on environmental issues: insights into public perception and its impact on low-carbon economy transition”. In: *Frontiers in Environmental Science* 12 (2024). DOI: 10.3389/fenvs.2024.1360304.

[31] T. Calamai et al. “Corporate greenwashing detection in text – a survey”. In: *arXiv preprint arXiv:2502.07541* (2025). URL: https://arxiv.org/abs/2502.07541.

[32] M. Capetz et al. “Leveraging Language Models to Detect Greenwashing”. In: *arXiv* (2023). DOI: 10.48550/arXiv.2311.01469.

[33] CDP. *CDP – Driving corporate transparency and environmental action.* URL: https://www.cdp.net/en.

[34] CDP. *CDP climate change questionnaire responses. Data from 01-01-22 to 31-12-22.* Data set. 2022.

[35] I. Cero, J. Luo, and J. M. Falligant. "Lexicon-Based Sentiment Analysis in Behavioral Research". In: *Perspectives on Behavior Science* 47.1 (2024), pp. 283–310. DOI: 10.1007/s40614-023-00394-x.

[36] S. Chandra, A. Kumar, and R. Singh. "A data-driven penalty-reward methodology for performance assessment of risk control systems". In: *Safety Science* 146 (2022), p. 105533. URL: https://www.sciencedirect.com/science/article/pii/S095042302200033X.

[37] J. N. Chauvey et al. "A study of the determinants of environmental disclosure quality: Evidence from French listed companies". In: *Journal of Management and Governance* 23.4 (2014), pp. 939–971. URL: https://link.springer.com/article/10.1007/s10997-019-09474-0.

[38] Y. Chen, X. Li, and H. Wang. "Can environmental information disclosure enhance firm value?—An analysis based on textual characteristics of annual reports". In: *PLOS ONE* 18.3 (2023), e0282394. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10001672/.

[39] C. H. Cho, R. W. Roberts, and D. M. Patten. "The language of US corporate environmental disclosure". In: *Accounting, Organizations and Society* 35.4 (2010), pp. 431–443. DOI: 10.1016/j.aos.2009.10.002.

[40] B. Choi, D. Lee, and J. Psaros. "An analysis of Australian company carbon emission disclosures". In: *Pacific Accounting Review* 25.1 (2013), pp. 58–79. DOI: 10.1108/01140581311318968.

[41] D. Choi, C. Y. Chung, and J. Young. "Analyzing and visualizing text information in corporate sustainability reports using natural language processing methods". In: *Applied Sciences* 12.11 (2019), p. 5614. URL: https://www.mdpi.com/2076-3417/12/11/5614.

[42] H. B. Christensen, L. Hail, and C. Leuz. "Mandatory CSR and sustainability reporting: Economic analysis and literature review". In: *Review of Accounting Studies* 26.3 (2021), pp. 1176–1248. URL: https://link.springer.com/article/10.1007/s11142-021-09609-5.

[43] P. M. Clarkson et al. "Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis". In: *Accounting, Organizations and Society* 33.4-5 (2008), pp. 303–327.

[44] Columbia Journalism Review. *In Spain, energy company buys covers of newspapers ahead of climate conference.* Dec. 2019. URL: https://www.cjr.org/business_of_news/endesa-climate-conference.php.

[45] CorpWatch. *Endesa accused of "academic greenwashing" at universities in Spain.* Nov. 2021. URL: https://www.corpwatch.org/article/endesa-accused-academic-greenwashing-universities-spain.

[46] G. D'Amore, M. Testa, and L. Lepore. "How Is the Utilities Sector Contributing to Building a Sustainable Future? A Systematic Literature Review of Sustainability Practices". In: *Sustainability* 16.1 (2023), p. 374. DOI: 10.3390/su16010374.

[47] Danwatch. *It makes up 75 percent of Denmark's "green" energy. Now it must be decided whether it can be called sustainable [Danish article].* Sept. 2023. URL: https://danwatch.dk/det-udgoer-75-procent-af-danmarks-groenne-energi-nu-skal-det-afgoeres-om-man-maa-kalde-det-baeredygtigt/.

[48] M. A. Delmas and V. C. Burbano. "The drivers of greenwashing". In: *California Management Review* 54.1 (2011), pp. 64–87. URL: https://journals.sagepub.com/doi/abs/10.1525/cmr.2011.54.1.64.

[49] M. A. Delmas and M. W. Toffel. "Organizational responses to environmental demands: Opening the black box". In: *Strategic Management Journal* 29.10 (2008), pp. 1027–1055. DOI: 10.1002/smj.701.

[50] Demagog. *Manipulation and greenwashing: We reveal the background of PGE's actions [Polish article]*. Oct. 2023. URL: https://demagog.org.pl/analizy_i_raporty/manipulacja-i-greenwashing-ujawniamy-kulisy-dzialan-pge/.

[51] M. Demartini, S. Trucco, and F. Testa. "Does the transparency of sustainability reports matter? A quantitative assessment". In: *Corporate Social Responsibility and Environmental Management* 32.2 (2025), p. 2926. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/csr.2926.

[52] G. S. Dhanesh and E. Nekmat. "Facts over stories for involved publics: Framing effects in CSR messaging and the roles of issue involvement, message elaboration, affect, and skepticism". In: *Management Communication Quarterly* 32.1 (2018), pp. 1–32. DOI: 10.1177/0893318918793941.

[53] Y. Ding and X. Liu. "A comparative analysis of data-driven methods in building energy benchmarking". In: *Energy and Buildings* 209 (2020), p. 109711. DOI: 10.1016/j.enbuild.2019.109711.

[54] M. P. Dočekalová, A. Kocmanová, and J. Hornungová. "Measuring and benchmarking corporate environmental performance". In: *Environmental software systems. Infrastructures, services and applications*. Ed. by R. Denzer et al. Springer, 2015, pp. 469–478. URL: https://link.springer.com/chapter/10.1007/978-3-319-15994-2_51.

[55] Y. Dou, J. Christensen, and H. Bruun. "Raising the bar: What determines the ambition level of corporate climate targets?" In: *Climatic Change* 167 (2024), p. 78. URL: https://link.springer.com/article/10.1007/s10584-024-03783-6.

[56] EEA. *Climate change - driving forces*. Figure. 2022. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Climate_change_-_driving_forces.

[57] P. S. Ellen, J. L. Wiener, and C. Cobb-Walgren. "The role of perceived consumer effectiveness in motivating environmentally conscious behaviors". In: *Journal of Public Policy & Marketing* 19.1 (2000), pp. 102–117. DOI: 10.1509/jppm.19.1.102.16949.

[58] Energiaklub. *EU taxonomy: Greenwashing impacts global climate crisis*. May 2022. URL: https://energiaklub.hu/en/article/eu-taxonomy-greenwashing-impacts-global-climate-crisis-5167.

[59] Energy Supply. *HOFOR and Ørsted reported to the consumer ombudsman for greenwashing biomass [Danish article]*. Mar. 2023. URL: https://www.energy-supply.dk/article/view/903005/hofor_og_%C3%98rsted_meldes_til_forbrugerombudsmanden_for_at_greenwashe_biomasse.

[60] ERG S.p.A. *Annual Report 2022*. Tech. rep. 2022. URL: https://www.annualreports.com/HostedData/AnnualReportArchive/e/erg-spa_2022.pdf.

[61] ERG S.p.A. *The Board of Directors of ERG approves the 2021–2025 Industrial Plan and ESG Plan*. Press release. May 2021. URL: https://www.erg.eu/en/-/il-consiglio-di-amministrazione-di-erg-approva-il-piano-industriale-e-il-piano-esg-2021-2025.

[62] C. Espinosa-Méndez and M. Jara-Bertín. "Mapping the greenwashing research landscape: A theoretical and field analysis". In: *Review of Managerial Science* (2025). URL: https://link.springer.com/article/10.1007/s11846-025-00856-3.

[63] Euronews. *German court dismisses lawsuit against RWE but makes 'landmark' ruling for future climate cases*. May 2025. URL: https://www.euronews.com/green/2025/05/28/the-mountains-have-won-german-court-dismisses-climate-suit-against-rwe-but-leaves-the-door.

[64] European Commission. *Composite indicators - Step 6: Weighting*. 2020. URL: https://knowledge4policy.ec.europa.eu/composite-indicators/10-step-guide/step-6-weighting_en.

[65] European Commission. *Corporate sustainability reporting.* 2024. URL: https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en.

[66] European Commission. *Directive (EU) 2024/825 of the European Parliament and of the Council of 28 February 2024 amending Directives 2005/29/EC and 2011/83/EU.* 2024. URL: https://eur-lex.europa.eu/eli/dir/2024/825/oj.

[67] European Commission. *Green claims.* 2023. URL: https://environment.ec.europa.eu/topics/circular-economy/green-claims_en.

[68] European Commission. *Guidance on the implementation/application of directive 2005/29/ec on unfair commercial practices.* Commission Staff Working Document SEC(2009) 1666 final. 2009. URL: https://www.eumonitor.eu/9353000/1/j4nvhdfdk3hydzq_j9vvik7m1c3gyxp/vio0dsm6rgzn.

[69] European Commission. *Proposal for a directive of the European Parliament and of the Council on common rules for the internal market in electricity.* COM(2016) 864 final. 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2016:0864:FIN.

[70] European Commission. *Proposal for a Directive of the European Parliament and of the Council on substantiation and communication of explicit environmental claims (Green Claims Directive).* COM(2023) 166 final. 2023. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52023PC0166.

[71] European Commission. *Proposal for a Directive on substantiation and communication of explicit environmental claims (Green Claims Directive). Recital 10.* Retrieved from EU public documents. 2024.

[72] European Commission. *Screening of websites for 'greenwashing': Half of green claims lack evidence.* Press release. Jan. 2021. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_269.

[73] European Parliament. *EU to ban greenwashing and improve consumer information on product durability.* Press release. Sept. 2023. URL: https://www.europarl.europa.eu/news/en/press-room/20230918IPR05412/eu-to-ban-greenwashing-and-improve-consumer-information-on-product-durability.

[74] European Union Agency for Fundamental Rights. *Enforcing consumer rights to combat greenwashing.* 2024. URL: https://fra.europa.eu/en/publication/2024/enforcing-consumer-rights-combat-greenwashing.

[75] J. Fan and R. Li. "Penalized estimating functions and variable selection in semiparametric regression models". In: *Journal of the American Statistical Association* 96.456 (2001), pp. 1348–1360. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2850080/.

[76] Federal Trade Commission. *Green Guides: Environmental marketing guidelines (16 C.F.R. § 260).* 2012. URL: https://www.ftc.gov/news-events/news/press-releases/2012/10/ftc-issues-revised-green-guides.

[77] K. Feghali, R. Najem, and B. D. Metcalfe. "Greenwashing in the era of sustainability: A systematic literature review". In: *Corporate Governance and Sustainability Review* 9.1 (2025), pp. 18–31. DOI: 10.22495/cgsrv9i1p2.

[78] T. Fiedler et al. "Assessments of the environmental performance of global companies need to account for company size". In: *Communications Earth & Environment* 5 (2024), p. 200. URL: https://www.nature.com/articles/s43247-024-01200-5.

[79] C. Fischer and F. Schoch. "Comparing human and NLP-based coding of interview transcripts". In: *Journal of Qualitative Social Research* 23.2 (2022), pp. 1–20.

[80] P. Folan and J. Browne. “A review of performance measurement: Towards performance management”. In: *Computers in Industry* 56.7 (2005), pp. 663–680. URL: https://www.sciencedirect.com/science/article/abs/pii/S0166361505000412.

[81] C. Free, S. Jones, and M.-S. Tremblay. “Greenwashing and sustainability assurance: A review and call for future research”. In: *Accounting, Governance and Regulation* (2024). Review. DOI: 10.XXXX/greenwashing-review.

[82] R. E. Freeman. *Strategic management: A stakeholder approach.* Pitman, 1984.

[83] A. A. Freitas. “Comprehensible classification models: A position paper”. In: *ACM SIGKDD Explorations Newsletter* 15.1 (2014), pp. 1–10. DOI: 10.1145/2594473.2594475.

[84] S. V. de Freitas Netto et al. “Concepts and forms of greenwashing: A systematic review”. In: *Environmental Sciences Europe* 32.1 (2020), p. 19. DOI: 10.1186/s12302-020-0300-3.

[85] J. Furnkranz, D. Gamberger, and N. Lavrač. *Foundations of rule learning.* Springer, 2018. DOI: 10.1007/978-3-662-55658-4.

[86] FXTOP. *Historical currency converter.* Retrieved July 21, 2025. URL: https://fxtop.com/en/historical-currency-converter.php.

[87] I. M. García-Sánchez et al. “Insights into sustainability reporting: Trends, aspects, and theoretical perspectives from a qualitative lens”. In: *Journal of Risk and Financial Management* 17.2 (2023), p. 68. DOI: 10.3390/jrfm17020068.

[88] L. Gatti, P. Seele, and L. Rademacher. “Grey zones in corporate sustainability communication: Taking stock of the credibility gap”. In: *International Journal of Corporate Social Responsibility* 6.1 (2021), pp. 1–16. DOI: 10.1186/s40991-021-00062-w.

[89] U. Golob, K. Podnar, and V. Zabkar. “Sustainability communication”. In: *International Journal of Advertising* 42.1 (2023), pp. 1–52. DOI: 10.1080/02650487.2022.2144035.

[90] S. Gourevitch and G. Gorham. *Carbon Disclosure Project Report 2008 – Global 500.* Tech. rep. UC San Diego, 2008. URL: https://gps.ucsd.edu/.../gorham.pdf.

[91] J. A. Grand-Guillaume-Perrenoud et al. “Mixed methods instrument validation: Evaluation procedures for practitioners”. In: *BMC Health Services Research* 23.1 (2023), Article 83. DOI: 10.1186/s12913-023-09040-3.

[92] S. Greco, M. Ehrgott, and J. R. Figueira. *Multiple criteria decision analysis: State of the art surveys.* Springer, 2016. URL: https://link.springer.com/book/10.1007/978-1-4939-3094-4.

[93] S. Greco et al. “On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness”. In: *Social Indicators Research* 141 (2018), pp. 61–94. URL: https://link.springer.com/article/10.1007/s11205-017-1832-9.

[94] Greenpeace Czech Republic. *Green Deal or greenwashing? Activists highlight problematic sponsors of major summit.* Oct. 2023. URL: https://www.greenpeace.org/czech/tiskova-zprava/19443/green-deal-or-greenwashing-activists-highlight-problematic-sponsors-of-major-summit/.

[95] Greenpeace Poland. *Lawsuit against PGE: Support Greenpeace’s legal defense [Polish article].* n.d. 2023. URL: https://dzialaj.greenpeace.pl/pozew.

[96] Greenpeace Poland. *PGE threatens Greenpeace with lawsuit over greenwashing campaign: "We won’t be silenced" [Polish article].* Oct. 2023. URL: https://www.greenpeace.org/poland/aktualnosci/31425/pge-grozi-greenpeace-pozwem-za-kampanie-demaskujaca-greenwashing-koncernu-aktywisci-nie-damy-sie-zastraszyc/.

[97] Greenpeace Poland. *PGE tries to intimidate Greenpeace [Polish article].* Oct. 2023. URL: https://www.greenpeace.org/poland/aktualnosci/31396/pge-probuje-zastraszyc-greenpeace/.

[98] J. Greer and K. Bruno. *Greenwash: The reality behind corporate environmentalism.* Third World Network, 1996.

[99] R. Guidotti et al. "A survey of methods for explaining black box models". In: *ACM Computing Surveys* 51.5 (2019), p. 93. DOI: 10.1145/3236009.

[100] R. Guo et al. "A path analysis of greenwashing in a trust crisis among Chinese energy companies: The role of brand legitimacy and brand loyalty". In: *Journal of Business Ethics* 140.3 (2017), pp. 523–536. DOI: 10.1007/s10551-015-2672-7.

[101] M. Hafeez et al. "Use of renewable energy to enhance firm performance: A systematic review". In: *Sustainability* 16.21 (2022), p. 9157. URL: https://www.mdpi.com/2071-1050/16/21/9157.

[102] T. Hale et al. "Assessing the rapidly emerging landscape of net zero targets". In: *Climate Policy* 22.1 (2022), pp. 18–29. DOI: 10.1080/14693062.2021.2013155.

[103] A. T. Hall et al. "Felt accountability in organizational performance: An emerging model and research agenda". In: *Journal of Organizational Behavior* 38.2 (2015), pp. 204–224. DOI: 10.1002/job.2052.

[104] S. Han, A. Wong, and T. Tan. "Expert-in-the-loop text categorization for sustainability reporting". In: *Computers in Industry* 146 (2023), p. 103820.

[105] P. S. Hart, S. N. Soroka, and D. Hiaeshutter Rice. "Go negative for clicks: Negative sentiment in environmental advocacy emails is associated with increased public engagement". In: *Environmental Communication* 18.3 (2024), pp. 287–304. DOI: 10.1080/17524032.2024.2372579.

[106] S. M. Hatefi and S. A. Torabi. "A common weight MCDA-DEA approach to construct composite indicators". In: *Ecological Economics* 70.1 (2010), pp. 114–120. URL: https://www.sciencedirect.com/science/article/abs/pii/S0921800910003381.

[107] S. A. Hazaea et al. "Sustainability assurance practices: A systematic review and future research agenda". In: *Environmental Science and Pollution Research* 29.4 (2022), pp. 4843–4865. URL: https://link.springer.com/article/10.1007/s11356-021-17359-9.

[108] N. Healy and J. Barry. "Politicizing energy justice and energy system transitions: Fossil fuel divestment and a "just transition"". In: *Energy Policy* 108 (2017), pp. 451–459. DOI: 10.1016/j.enpol.2017.06.014.

[109] S. Hill et al. "Testing the greenwashing assessment framework". In: *Ecology and Society* 30.2 (2025), Article 31. DOI: 10.5751/ES-16106-300231.

[110] J. Huang and H. Guo. "When a bleak future comes closer: interaction effects of emotion and temporal distance framing in climate change communication". In: *BMC Psychology* 12.1 (2024). DOI: 10.1186/s40359-024-02183-w.

[111] K. Hyland. "Talking to the academy: Forms of hedging in science research articles". In: *Written Communication* 13.2 (1996), pp. 251–281. DOI: 10.1177/0741088396013002004.

[112] Ø. Ihlen, J. Bartlett, and S. May. *The handbook of communication and corporate social responsibility.* Wiley-Blackwell, 2011. DOI: 10.1002/9781118087133.

[113] International Energy Agency. *Net Zero by 2050: A Roadmap for the Global Energy Sector*. Tech. rep. 2021. URL: https://www.iea.org/reports/net-zero-by-2050.

[114] International Hydropower Association. *Scatec ASA*. Retrieved July 24, 2025. 2025. URL: https://www.hydropower.org/our-members/scatec.

[115] K. Isaacs et al. "Set ambitious but realistic environmental goals". In: *MIT Sloan Management Review* (Dec. 2023). URL: https://sloanreview-mit-edu.tudelft.idm.oclc.org/article/set-ambitious-but-realistic-environmental-goals/.

[116] I. Jaffari, R. Peterson, and D. Lee. "Textual analysis of ESG disclosures and their impact on ESG ratings". In: *Finance Research Letters* 52 (2024), p. 103478.

[117] M. D. T. de Jong, G. Huluba, and A. D. Beldad. “Different shades of greenwashing: Consumers’ reactions to environmental lies, half-lies, and legal compliance”. In: *Journal of Business Research* 107 (2020), pp. 36–49. DOI: 10.1016/j.jbusres.2018.09.009.

[118] S. Jørgensen et al. “Limited accountability and awareness of corporate emissions target outcomes”. In: *Nature Climate Change* 14 (2024), p. 236. URL: https://www.nature.com/articles/s41558-024-02236-3.

[119] T. J. Kaitharath. “Greenwashing - A marketing tactic or genuine commitment?” In: *Journal of Management Research and Analysis* 10.1 (2023), pp. 1–2. DOI: 10.18231/j.jmra.2023.001.

[120] H. Kang and J. Kim. “Analyzing and visualizing text information in corporate sustainability reports using natural language processing methods”. In: *Applied Sciences* 12.11 (2022), p. 5614. DOI: 10.3390/app12115614.

[121] H. Kang and J. Kim. “Analyzing and visualizing text information in corporate sustainability reports using natural language processing methods”. In: *Applied Sciences* 12.11 (2022), p. 5614. DOI: 10.3390/app12115614.

[122] R. Kashyap, S. Verma, and R. Pal. “ESG disclosures and firm value: Evidence from India”. In: *Journal of Indian Business Research* 13.2 (2020), pp. 234–253. DOI: 10.1108/JIBR-11-2019-0340.

[123] M. C. Kathan et al. “What you see is not what you get: ESG scores and greenwashing risk”. In: *Finance Research Letters* 74 (2025), p. 106710. DOI: 10.1016/j.frl.2024.106710.

[124] R. Kemp, E. Pontikes, and O. Hawn. “Set ambitious but realistic environmental goals”. In: *MIT Sloan Management Review* 65.3 (2024), p. 23. URL: https://sloanreview.mit.edu/article/set-ambitious-but-realistic-environmental-goals/.

[125] E.-H. Kim and T. P. Lyon. “Greenwash vs. brownwash: Exaggeration and undue modesty in corporate sustainability disclosure”. In: *Organization Science* 26.3 (2015), pp. 705–723. DOI: 10.1287/orsc.2014.0949.

[126] K. Kluza. “Elimination of sandbagging behaviour in budgeting process for short term and long term goal setting”. In: *Proceedings of the 33rd International Business Information Management Association Conference (IBIMA)*. 2019. URL: http://hdl.handle.net/20.500.12182/819.

[127] T. Koellner et al. “A novel multilevel decision-making evaluation approach for the renewable energy heating systems: A case study in China”. In: *Journal of Cleaner Production* 423 (2023), p. 926. URL: https://www.sciencedirect.com/science/article/abs/pii/S0959652623000926.

[128] U. Kosi and P. Relard. “Are firms (getting) ready for the corporate sustainability reporting directive?” In: *Sustainability Accounting, Management and Policy Journal* (2024). DOI: 10.1007/s00550-024-00541-1.

[129] S. K. Kwak and J. H. Kim. “Statistical data preparation: Management of missing values and outliers”. In: *Korean Journal of Anesthesiology* 70.4 (2017), pp. 407–411. DOI: 10.4097/kjae.2017.70.4.407.

[130] V. Lagasio. “ESG-washing detection in corporate sustainability reports”. In: *International Review of Financial Analysis* 96 (2024), p. 103742. DOI: 10.1016/j.irfa.2024.103742.

[131] E. Lähdesmäki and M. Mattinen. “Sustainability in energy companies under the lens of cultural framing: Genuine transformation or rebranding strategy?” In: *Corporate Social Responsibility and Environmental Management* (2024). DOI: 10.1002/csr.3111.

[132] H. Li, P. Guo, and S. Huang. “Third-party environmental information disclosure and firms’ carbon emissions”. In: *Research in International Business and Finance* 58 (2021), p. 101486. URL: https://www.sciencedirect.com/science/article/abs/pii/S0140988324000586.

[133] S. Li and G. Tang. “Stakeholder pressures and corporate environmental strategies: Evidence from China’s listed firms”. In: *Sustainability* 12.3 (2020), p. 1172. DOI: 10.3390/su12031172.

[134] X. Li et al. "Electrification and digitalization effects on sectoral energy demand and consumption: A prospective study towards 2050". In: *Energy* 279 (2023), p. 127992. DOI: 10.1016/j.energy.2023.127992.

[135] Y. Li, X. Wang, and H. Chen. "Sentiment analysis of global news on environmental issues: Insights into public perception and its impact on low-carbon economy transition". In: *Frontiers in Environmental Science* 12 (2024), p. 1360304. URL: https://www.frontiersin.org/articles/10.3389/fenvs.2024.1360304/full.

[136] Z. Li and L. Chen. "A corpus-based quantitative study on the interpersonal functions of hedges in Chinese and German academic discourse". In: *Frontiers in Psychology* 13 (2022), p. 959555. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.959555/full.

[137] H. Liang et al. "Future-Time Framing: The Effect of Language on Corporate Future Orientation". In: *Organization Science* 29.6 (2018), pp. 1093–1111. DOI: 10.1287/orsc.2018.1217.

[138] LSEG. *Environmental, social and governance scores from LSEG: Methodology*. 2024. URL: https://www.lseg.com/content/dam/data-analytics/en_us/documents/methodology/lseg-esg-scores-methodology.pdf.

[139] N. H. Luu et al. "Does mandatory greenhouse gas emissions reporting program deter corporate greenwashing?" In: *Journal of Environmental Management* 373 (2025), p. 123740. DOI: 10.1016/j.jenvman.2024.123740.

[140] T. P. Lyon and J. W. Maxwell. "Greenwash: Corporate environmental disclosure under threat of audit". In: *Journal of Economics & Management Strategy* 20.1 (2011), pp. 3–41. DOI: 10.1111/j.1530-9134.2010.00282.x.

[141] T. P. Lyon and A. W. Montgomery. "The means and end of greenwash". In: *Organization & Environment* 28.2 (2015), pp. 223–249. URL: https://journals.sagepub.com/doi/10.1177/1086026615575332.

[142] M. Malik and O. Terzidis. "Hybrid frameworks in AI-augmented literature reviews: Ensuring rigor and transparency". In: *International Journal of Research Methodology* 18.1 (2025), pp. 1–15.

[143] H. B. Mann and D. R. Whitney. "On a test of whether one of two random variables is stochastically larger than the other". In: *The Annals of Mathematical Statistics* 18.1 (1947), pp. 50–60. DOI: 10.1214/aoms/1177730491.

[144] F. Mariani and M. Ciommi. "Aggregating Composite Indicators through the Geometric Mean: A Penalization Approach". In: *Computation* 10.4 (2022), p. 64. DOI: 10.3390/computation10040064.

[145] M. Marijnissen. *CSRD regulation timeline*. Retrieved July 10, 2025. URL: https://intire.nl/csrd-2022-delays/.

[146] MarketingNews. *Endesa accused of 'greenwashing' for its sponsorship of COP25 [Spanish article]*. Dec. 2019. URL: https://www.marketingnews.es/marcas/noticia/1156298054305/endesa-acusada-de-greenwashing-patrocinio-de-cop25.1.html.

[147] Marketscreener. *Atlantica Sustainable Infrastructure Plc company*. Marketscreener. Retrieved July 21, 2025. URL: https://www.marketscreener.com/quote/stock/ATLANTICA-SUSTAINABLE-INF-16686514/company/.

[148] Marketscreener. *E.D.F. company*. Marketscreener. Retrieved July 21, 2025. URL: https://www.marketscreener.com/quote/stock/E-D-F-4998/company/.

[149] C. Marquis, M. W. Toffel, and Y. Zhou. "Scrutiny, Norms, and Selective Disclosure: A Global Study of Greenwashing". In: *Organization Science* 27.2 (2016), pp. 483–504. DOI: 10.1287/orsc.2015.1039.

[150] A. J. Mateo-Márquez, J. M. González-González, and C. Zamora-Ramírez. “An international empirical study of greenwashing and voluntary carbon disclosure”. In: *Journal of Cleaner Production* 363 (2022), p. 132567. DOI: 10.1016/j.jclepro.2022.132567.

[151] D. H. Matthews. “Environmental management systems for internal corporate environmental benchmarking”. In: *Benchmarking: An International Journal* 10.2 (2003), pp. 95–106. URL: https://www.emerald.com/insight/content/doi/10.1108/14635770310469635/full/html.

[152] J. Mayer, H. Chung, and V. Rao. “In the loop: Refining AI outputs in domain-specific text analysis”. In: *Journal of Applied Artificial Intelligence* 37.4 (2023), pp. 233–246.

[153] D. Mehta. “Public perception of corporate environmental commitments: How communication shapes trust”. In: *Journal of Law for Future Security* 1.2 (2024), pp. 63–74. DOI: 10.54688/jlfs.v1i2.452.

[154] S. Mena and S. Parker. “The temporal structuring of corporate sustainability”. In: *Journal of Business Ethics* 195.1 (2024), pp. 1–23. DOI: 10.1007/s10551-024-05648-5.

[155] J. W. Meyer and B. Rowan. “Institutionalized organizations: Formal structure as myth and ceremony”. In: *American Journal of Sociology* 83.2 (1977), pp. 340–363. DOI: 10.1086/226550.

[156] G. Michelon, S. Pilonato, and F. Ricceri. “CSR reporting practices and the quality of disclosure: An empirical analysis”. In: *Critical Perspectives on Accounting* 33 (2014), pp. 59–78. DOI: 10.1016/j.cpa.2014.10.003.

[157] J. Millek. “The robustness of TOPSIS results using sensitivity analysis based on weight tuning”. In: *World Congress on Medical Physics and Biomedical Engineering 2018 (IFMBE Proceedings, Vol. 68/2)*. Ed. by L. Lhotska et al. Springer, 2018, pp. 1009–1016. DOI: 10.1007/978-981-10-9038-7_15.

[158] J. Mingers. “Combining IS research methods: Towards a pluralist methodology”. In: *Information Systems Research* 12.3 (2001), pp. 240–259. DOI: 10.1287/isre.12.3.240.9709.

[159] A. W. Montgomery, T. P. Lyon, and J. Barg. “No end in sight? A greenwash review and research agenda”. In: *Organization & Environment* 37.2 (2024), pp. 221–256. DOI: 10.1177/10860266231168905.

[160] Morningstar. *Arendals Fossekompani ASA quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/xosl/afk/quote.

[161] Morningstar. *CEZ AS quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/pinx/czavf/quote.

[162] Morningstar. *EDP - Energias de Portugal S.A. quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/pinx/elcpf/quote.

[163] Morningstar. *Endesa S.A. quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/pinx/elezf/quote.

[164] Morningstar. *ERG S.p.A. quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/xmil/erg/quote.

[165] Morningstar. *Ørsted A/S quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/pinx/dogef/quote.

[166] Morningstar. *Polska Grupa Energetyczna S.A. quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/xwar/pge/quote.

[167] Morningstar. *Romande Energie Holding SA quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/xswx/rehn/quote.

[168] Morningstar. *Scatec ASA quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/pinx/stecf/quote.

[169] Morningstar. *Solaria Energia y Medio Ambiente S.A. quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/pinx/seymf/quote.

[170] Morningstar. *Terna Energy S.A. quote*. Morningstar. Retrieved July 21, 2025. URL: https://www.morningstar.com/stocks/pinx/treaf/quote.

[171] N. Nachar. "The Mann–Whitney U: A test for assessing whether two independent samples come from the same distribution". In: *Tutorials in Quantitative Methods for Psychology* 4.1 (2008), pp. 13–20. DOI: 10.20982/tqmp.04.1.p013.

[172] N. Nascimento et al. "LLM4DS: Evaluating Large Language Models for Data Science Code Generation". In: *arXiv* (2024). URL: https://arxiv.org/abs/2411.11908.

[173] N. Nemes et al. "An integrated framework to assess greenwashing". In: *Sustainability* 14.8 (2022), p. 4431. DOI: 10.3390/su14084431.

[174] Network for Greening the Financial System. *Information note on improving GHG emissions data*. Tech. rep. 2024. URL: https://www.ngfs.net/sites/default/files/medias/documents/ngfs_information_note_on_improving_ghg_emission_data.pdf.

[175] M. Nipper, A. Ostermaier, and J. Theis. "Mandatory disclosure of standardized sustainability metrics: The case of the EU Taxonomy Regulation". In: *arXiv* (2022). DOI: 10.48550/arXiv.2205.15576.

[176] B. O'Dwyer and D. L. Owen. "External transparency or internal capture? The role of third-party statements in adding value to corporate environmental reports". In: *Business Strategy and the Environment* 9.1 (2000), pp. 1–23. DOI: 10.1002/(SICI)1099-0836(200001/02)9:1<1::AID-BSE227>3.0.CO;2-H.

[177] M. Odobaša and Ž. Marošević. *Insights into companies' adoption of the Corporate Sustainability Reporting Directive: Historical, legal, and economic context*. 2023. URL: https://www.diva-portal.org/smash/get/diva2:1873308/FULLTEXT02.pdf.

[178] OECD. *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing, 2008. URL: https://www.oecd.org/sdd/42495745.pdf.

[179] OneTrust. *EFRAG issues draft European sustainability reporting standards*. 2022. URL: https://www.onetrust.com/blog/efrag-eu-sustainability-reporting-standards/.

[180] E. Ortiz-Martínez and S. Marín-Hernández. "Sustainability Information in European Small- and Medium-Sized Enterprises". In: *Journal of the Knowledge Economy* 15.2 (2023), pp. 7497–7522. DOI: 10.1007/s13132-023-01386-7.

[181] J. W. Osborne and A. Overbay. "The power of outliers (and why researchers should always check for them)". In: *Practical Assessment, Research & Evaluation* 9.6 (2004). URL: http://pareonline.net/getvn.asp?v=9&n=6.

[182] Oxford Net Zero. *Getting net zero right*. 2021. URL: https://netzeroclimate.org/what-is-net-zero/getting-net-zero-right/.

[183] S. Papadaki et al. "Evaluation of renewable energy system for sustainable development: A methodology for sustainable development". In: *Renewable Energy and Environmental Sustainability* 6 (2021), p. 41. URL: https://www.rees-journal.org/articles/rees/full_html/2021/01/rees210041/rees210041.html.

[184] A. Pathania and M. Rastogi. "Greening the Corporate Landscape: the Role of CSR in Environmental Sustainability". In: *International Journal For Multidisciplinary Research* 6.3 (2024). DOI: 10.36948/ijfmr.2024.v06i03.22639.

[185] S. Pervaiz, G. Li, and Q. He. "The mechanism of goal-setting participation's impact on employees' proactive behavior: Moderated mediation role of power distance". In: *PLOS ONE* 16.12 (2021), e0260625. DOI: 10.1371/journal.pone.0260625.

[186] R. G. Peters, V. T. Covello, and D. B. McCallum. "The determinants of trust and credibility in environmental risk communication: An empirical study". In: *Risk Analysis* 17.1 (1997), pp. 43–54. DOI: 10.1111/j.1539-6924.1997.tb00842.x.

[187] T. Pimonenko et al. "Green brand of companies and greenwashing under sustainable development goals". In: *Sustainability* 12.4 (2020), p. 1679. DOI: 10.3390/su12041679.

[188] D. Plaehn. "What's the real penalty in penalty analysis?" In: *Food Quality and Preference* 28.1 (2013), pp. 323–329. DOI: 10.1016/j.foodqual.2012.11.009.

[189] Polska Grupa Energetyczna. *Integrated Report 2022*. Tech. rep. 2022. URL: https://raportzintegrowany2022.gkpge.pl/en/wp-content/uploads/sites/2/2023/12/PGEskrotraportuEN.pdf.

[190] R. Pramudita, M. Johansson, and S. Wang. "The evolving landscape of CSR report analysis: Moving beyond manual coding". In: *Sustainability Accounting, Management and Policy Journal* 14.2 (2023), pp. 202–220.

[191] W. Przepiorka and C. Horne. "How can consumer trust in energy utilities be increased? The effectiveness of prosocial, proenvironmental, and service-oriented investments as signals of trustworthiness". In: *Organization & Environment* 33.2 (2020), pp. 262–284. DOI: 10.1177/1086026618803729.

[192] M. S. Rahman et al. "Review and evaluation of performance measures for survival prediction models in external validation settings". In: *BMC Medical Research Methodology* 17 (2017), Article 60. DOI: 10.1186/s12874-017-0336-2.

[193] Refinitiv. *Refinitiv ESG scores methodology guide*. Tech. rep. London Stock Exchange Group, 2022. URL: https://www.lseg.com/sites/default/files/content/documents/Refinitiv%20ESG%20Scores%20Methodology.pdf.

[194] *Regulation (EU) 2020/852 of the European Parliament and of the Council of 18 June 2020 on the establishment of a framework to facilitate sustainable investment (EU Taxonomy Regulation)*. 2020. URL: https://eur-lex.europa.eu/eli/reg/2020/852/oj.

[195] D. L. Reichmuth, L. Ford, and J. Hierholzer. "Harmonizing corporate carbon footprints: Accounting for reporting inconsistency, boundary incompleteness, and activity exclusion". In: *Nature Communications* 12 (2021), p. 2400. DOI: 10.1038/s41467-021-22730-x.

[196] ReportYak. *Greenwashing: Uncovering misleading environmental claims*. 2024. URL: https://reportyak.com/blog/understanding-greenwashing/.

[197] Reuters. *Akenerji Elektrik Uretim AS company information*. Reuters. Retrieved July 21, 2025. URL: https://www.reuters.com/markets/companies/AKENR.IS/.

[198] S. G. Roch. "Absolute vs relative performance rating formats: Implications for fairness and organizational justice". In: *International Journal of Selection and Assessment* 15.4 (2007), pp. 390–399. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2389.2007.00390.x.

[199] C. Rudin. "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead". In: *Nature Machine Intelligence* 1.5 (2019), pp. 206–215. DOI: 10.1038/s42256-019-0048-x.

[200] J. L. Ruiz and I. Sirvent. "Performance evaluation through DEA benchmarking adjusted to goals". In: *Omega* 87 (2019), pp. 150–157. DOI: 10.1016/j.omega.2018.08.014.

[201] M. Saisana, A. Saltelli, and S. Tarantola. "Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168.2 (2005), pp. 307–323. URL: https://academic.oup.com/jrsssa/article/168/2/307/7084307.

[202] S. Saxena. "Using NLP for detecting greenwashing indicators and constructing impact-focused index portfolios". Master's thesis. University College London, 2024. DOI: 10.2139/ssrn.4739445.

[203] D. Schmuck, J. Matthes, and B. Naderer. “Misleading consumers with green advertising? An affect–reason–involvement account of greenwashing effects in environmental advertising”. In: *Journal of Advertising* 47.2 (2018), pp. 127–145. DOI: 10.1080/00913367.2018.1452652.

[204] A. K. Schnackenberg and E. C. Tomlinson. “Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships”. In: *Journal of Management* 42.7 (2016), pp. 1784–1810. DOI: 10.1177/0149206314525202.

[205] D. Schoenmaker, W. Schramade, and G. Spaargaren. “Harmonizing corporate carbon footprints”. In: *Nature Communications* 12 (2021), p. 6349. URL: https://www.nature.com/articles/s41467-021-26349-x.

[206] Science Based Targets initiative. *Science-based targets result in biggest emissions reduction to date*. 2022. URL: https://sciencebasedtargets.org/reports/sbti-progress-report-2021/science-based-targets-result-in-biggest-emissions-reduction-to-date.

[207] Science Based Targets initiative. *Understand the methods for science-based climate action*. 2024. URL: https://sciencebasedtargets.org/news/understand-science-based-targets-methods-climate-action.

[208] C. Seale. “Quality in qualitative research”. In: *Qualitative Inquiry* 5.4 (1999), pp. 465–478. DOI: 10.1177/107780049900500402.

[209] P. Seele and L. Gatti. “Greenwashing revisited: In search of a typology and accusation-based definition incorporating legitimacy strategies”. In: *Business Strategy and the Environment* 26.2 (2017), pp. 239–252. DOI: 10.1002/bse.1912.

[210] Z. Shen et al. “LayoutParser: A unified toolkit for deep learning based document image analysis”. In: *arXiv preprint arXiv:2103.15348* (2021). DOI: 10.48550/arXiv.2103.15348.

[211] T. Shimamura, Y. Tanaka, and S. Managi. “Decoding greenwashing: LLM insights into corporate narrative”. In: *SSRN Working Paper* (2025). DOI: 10.2139/ssrn.5253231.

[212] E. H. Shortliffe and M. J. Sepúlveda. “Clinical decision support in the era of artificial intelligence”. In: *JAMA* 320.21 (2018), pp. 2199–2200. DOI: 10.1001/jama.2018.17163.

[213] F. Simon and V. Schweitzer. “When smart meters backfire on energy transition internalization: Ethical electricity suppliers’ mitigation of consumer data vulnerability and attendant psychological disempowerment”. In: *Technological Forecasting and Social Change* 194 (2023), p. 122738. DOI: 10.1016/j.techfore.2023.122738.

[214] I. Sobiech, A. Kowalkowski, and M. Marcinkowska. “ESG analysis of companies included in the RESPECT Index based on Thomson Reuters EIKON database”. In: *Economic and Environmental Studies* 18.4 (2018), pp. 1603–1621. DOI: 10.15611/pn.2018.520.10.

[215] Sortir du nucléaire. *EDF, greenwashing in its purest form! [French article]*. May 2021. URL: https://www.sortirdunucleaire.org/EDF-le-greenwashing-au-naturel.

[216] B. K. Sovacool and S. E. Ryan. *The sociology of energy, buildings and the environment: Constructing knowledge, designing practice*. Routledge, 2016. DOI: 10.4324/9781315685995.

[217] spaCy Documentation. *Models & pipelines: en_core_web_lg*. Retrieved July 21, 2025. URL: https://spacy.io/models/en%5C#en_core_web_lg.

[218] spaCy Documentation. *Rule-based matching*. Retrieved July 21, 2025. URL: https://spacy.io/usage/rule-based-matching.

[219] M. J. Spaniol et al. “Defining greenwashing: A concept analysis”. In: *Sustainability* 16.20 (2024), Article 9055. DOI: 10.3390/su16209055.

[220] M. Spence. “Job market signaling”. In: *The Quarterly Journal of Economics* 87.3 (1973), pp. 355–374. DOI: 10.2307/1882010.

[221] M. B. Steinborn et al. "Methodology of performance scoring in the d2 sustained-attention test: Cumulative-reliability functions and practical guidelines". In: *Psychological Assessment* 29.12 (2017), pp. 1590–1603. DOI: 10.1037/pas0000482.

[222] W. Strielkowski et al. "Renewable energy in the sustainable development of electrical power sector: A review". In: *Energies* 14.24 (2021), p. 8240. DOI: 10.3390/en14248240.

[223] R. Su et al. "Strategies for building environmental transparency and accountability". In: *Sustainability* 13.16 (2021), p. 9116. URL: https://www.mdpi.com/2071-1050/13/16/9116.

[224] Sustainalytics. *Atlantica Sustainable Infrastructure PLC ESG risk rating*. Retrieved July 24, 2025. 2025. URL: https://www.sustainalytics.com/esg-rating/atlantica-sustainable-infrastructure-plc/1269194327.

[225] P. Tashman, V. Marano, and T. Kostova. "Walking the walk or talking the talk? Corporate social responsibility decoupling in emerging market multinationals". In: *Journal of International Business Studies* 50.2 (2019), pp. 153–171. DOI: 10.1057/s41267-018-0171-7.

[226] T. Tays, P. Johnson, and A. Gamarra. *Environmental performance metrics in incentive plans: Incentive trends and key design considerations*. July 2024. URL: https://corpgov.law.harvard.edu/2024/07/23/environmental-performance-metrics-in-incentive-plans-incentive-trends-and-key-design-considerations/.

[227] F. Testa et al. "Does it pay to be a greenwasher or a brownwasher?" In: *Business Strategy and the Environment* 27.7 (2018), pp. 1104–1116. DOI: 10.1002/bse.2058.

[228] R. Torelli, F. Balluchi, and K. Furlotti. "Greenwashing and environmental communication: Effects on stakeholders' perceptions". In: *Business Strategy and the Environment* 29.2 (2020), pp. 407–421. DOI: 10.1002/bse.2373.

[229] D. Trilling and J. G. Jonkman. "Scaling up content analysis". In: *Communication Methods and Measures* 12.2–3 (2018), pp. 158–174.

[230] UN High-Level Expert Group on Net Zero Emissions Commitments. *Integrity matters: Net zero commitments by businesses, financial institutions, cities and regions*. Tech. rep. United Nations, 2022. URL: https://www.un.org/sites/un2.un.org/files/high-level_expert_group_n7_final_report.pdf.

[231] United Nations. *Greenwashing – the deceptive tactics behind environmental claims*. 2024. URL: https://www.un.org/en/climatechange/science/climate-issues/greenwashing.

[232] Utility Dive. *The clean energy transition is here — and so are greenwashing claims*. 2023. URL: https://www.utilitydive.com/news/clean-energy-transition-environmental-greenwashing-claims-consumer-protection/702668/.

[233] B. Van der Kolk, H. J. ter Bogt, and P. M. van Veen-Dirks. "The use of performance measurement systems in the public sector: Effects on performance". In: *Management Accounting Research* 46 (2019), p. 693. URL: https://www.sciencedirect.com/science/article/abs/pii/S1044500513000693.

[234] S. F. Verde et al. "Comparing the ambition of EU companies with science-based targets to EU regulation-imposed reductions". In: *npj Climate Action* 3 (2024), p. 98. URL: https://www.nature.com/articles/s44168-024-00098-1.

[235] A. Vinella et al. "Leveraging language models to detect greenwashing". In: *arXiv preprint arXiv:2311.01469* (2023). URL: https://arxiv.org/abs/2311.01469.

[236] K. Walker and F. Wan. "The harm of symbolic actions and green-washing: Corporate actions and communications on environmental performance and their financial implications". In: *Journal of Business Ethics* 109.2 (2012), pp. 227–242. DOI: 10.1007/s10551-011-1122-4.

[237] C. Wang, P. Chen, and Y. Hao. "Tax incentives and green innovation—The mediating role of financing constraints and the moderating role of subsidies". In: *Frontiers in Environmental Science* 10 (2022), p. 1067534. URL: https://www.frontiersin.org/articles/10.3389/fenvs.2022.1067534/full.

[238] X. Wang, X. Gao, and M. Sun. "Construction and analysis of corporate greenwashing index: A deep learning approach". In: *EPJ Data Science* 14 (2025), p. 44. DOI: 10.1140/epjds/s13688-025-00562-w.

[239] X. Wang et al. "Evaluation of provincial power supply reliability with high penetration of renewable energy based on combination weighting of game theory-TOPSIS method". In: *Applied Energy* 328 (2024), p. 120005. URL: https://www.sciencedirect.com/science/article/abs/pii/S2352467723001005.

[240] Z. Wang et al. "Textual attributes of corporate sustainability reports and ESG ratings". In: *Sustainability* 16.21 (2024), p. 9270. URL: https://www.mdpi.com/2071-1050/16/21/9270.

[241] A. Wasi. *Greenwashing exposed: Analyzing the impact on consumer trust, brand reputation, and the environment*. 2024. URL: https://medium.com/@aleezawasi/greenwashing-exposed-analyzing-the-impact-on-consumer-trust-brand-reputation-and-the-environment-9b955161995c.

[242] N. Webersinke et al. "ClimateBERT: A pretrained language model for climate-related text". In: *arXiv* (2021). DOI: 10.48550/arXiv.2110.12010.

[243] F. Wilcoxon. "Individual comparisons by ranking methods". In: *Biometrics Bulletin* 1.6 (1945), pp. 80–83. DOI: 10.2307/3001968.

[244] World Benchmarking Alliance. *Energias de Portugal – Electric utilities benchmark*. 2023. URL: https://www.worldbenchmarkingalliance.org/publication/electric-utilities/companies/energias-de-portugal/.

[245] World Economic Forum. *IEA: Clean energy investment must reach $4.5 trillion per year by 2030 to limit global warming to 1.5°C*. 2023. URL: https://www.weforum.org/stories/2023/09/iea-clean-energy-investment-global-warming/.

[246] P. Wright, J. Allen, and L. Zhao. "Bias and reproducibility issues in manual qualitative coding". In: *Qualitative Research Journal* 22.1 (2022), pp. 1–22.

[247] J. Xu. "A natural language processing based technique for sentiment analysis of college english corpus". In: *PeerJ Computer Science* 9 (2023), e1235. DOI: 10.7717/peerj-cs.1235.

[248] M. Yang, L. Wang, and X. Chen. "How does green communication promote the green consumption intention of social media users?" In: *Journal of Retailing and Consumer Services* 68 (2022), p. 103008. DOI: 10.1016/j.jretconser.2022.103008.

[249] Z. Y. Yeo, R. T. Ng, and O. Heidrich. "What drives companies' progress on their emission reduction targets?" In: *Journal of Cleaner Production* 464 (2024), p. 142733. URL: https://www.sciencedirect.com/science/article/pii/S0959652624025733.

[250] L. Zhang, M. Wang, and S. Liu. "Bridging the gap in ESG measurement: Using NLP to quantify environmental, social, and governance communication". In: *Journal of Banking & Finance* 160 (2024), p. 107086. URL: https://www.sciencedirect.com/science/article/pii/S1544612324000096.

[251] M. Zhao, L. Fernández, and P. Kumar. "Automated detection of greenwashing in corporate climate communication: A survey of NLP approaches". In: *Journal of Environmental Communication* 17.3 (2023), pp. 243–260.

[252] M. Ziółko. "Literature review of greenwashing research: State of the art". In: *Corporate Social Responsibility and Environmental Management* 31.3 (2024), pp. 1428–1445. DOI: 10.1002/csr.2456.

Appendices

A Company Sample Profiles

This Appendix gives information for each of the 14 companies within the final sample. Company profiles below draw from recent third-party business descriptions (Reuters, Morningstar, Marketscreener) supplemented with operational data from CDP 2022 submissions, noting that current business structures may differ from the 2021-2022 analysis period because of ongoing industry transformation.

Akenerji Elektrik Üretim A.Ş. is "a Turkey-based company engaged in the power and steam generation" that "also makes investments in the renewable energy sector" and "is a member of the Akkok Group of Companies" [197]. The company operates with "1,224 MW" installed capacity where "26% of our installed power...consists of renewable energy sources" across "1 natural gas combined cycle, 1 wind power plant and 7 hydroelectric power plants" [34].

Arendals Fossekompani ASA is "an industrial investment company holding several core investments" that "invests in and owns companies that make energy from renewable sources more usable and accessible" and "owns and operates two hydropower plants" [160]. "Established in 1896 to harness the energy from an everlasting natural resource; water," and the company "has been listed on the Oslo Stock Exchange since 1913" [34].

Atlantica Sustainable Infrastructure PLC is "a sustainable infrastructure company with a majority of its business in renewable energy assets" with facilities "in North America...South America...and EMEA" [147]. The company maintains "2,121 MW of aggregate renewable energy installed generation capacity (of which 73% is solar)" where "our renewable sector represented 75% of our revenue with solar energy representing 64%" [34].

CEZ is "a Czech energy company of which the government of the Czech Republic is the majority shareholder" with "core business...the generation, distribution, trade, and sale of electricity and heat" [161]. The utility operates "nuclear, coal-fired, gas-fired, hydroelectric, photovoltaic (PV), wind, and biogas facilities" and employs "more than 28,000 people" [34].

EDF (Électricité de France) is a "France-based company...predominantly engaged in carbon-free electricity in power plants" and functions as "Producer and distributor of gas and electricity for utility purposes to individual and commercial clients" [148]. The utility maintains "a global installed net generation capacity of 116.9 GWe" with "171,490 employees" serving "approximately 40.3 million customers" where "90% of the generation is carbon free at Group level" [34].

EDP (Energias de Portugal) is "a vertically integrated utility company and is the largest generator, supplier, and distributor of electricity in Portugal" that "owns 71% of EDP Renovaveis, the fourth-largest wind power owner/operator in the world" [162]. The company had "an installed capacity of around 26 Gigawatt (GW) (79% renewable)" with "around 13.2 thousand employees" and "generated about 61 Terawatt hour (TWh) of electricity worldwide, of which 74% from renewable energy sources" [34].

ERG S.p.A "produces electricity from clean, renewable, and sustainable sources" where "wind power production is handled by subsidiary ENG Renew" and operates "to markets across Europe" [164]. The company maintains "2,944 MW" installed capacity after completing "its transformation towards a pure 'Wind & Solar' business model" following asset divestments [34].

Endesa "generates, distributes, and supplies electricity in Spain and Portugal" with generation "split among hydroelectric, nuclear, natural gas, oil, solar, and wind" and "also supplies gas to retail and business customers in Spain and France" [163]. The utility maintains "a total power installed capacity of 22,044 MW" where "renewables represent 9,293 MW (approximately 42% of the total)" with "9,258 employees" [34].

PGE (Polska Grupa Energetyczna SA) is "a Polish electric utility company of which the Polish State Treasury is the majority shareholder" that "controls a portfolio of lignite, coal, gas, biomass, hydro, and wind power plants" [166]. The utility employs "almost 40,000 employees" and "accounts for approx. 40% of electricity output in Poland" covering "the entire value chain: from lignite mining, through generation of electricity and heat, to distribution and sales" [34].

Romande Energie Holding SA is "a Swiss utility company involved in the generation, distribution, and marketing of electricity and energy services" that "generates maximum revenue from the Energy Solutions segment" [167]. The utility employs "more than 1.250 employees working in 5 different cantons" where "all our generation assets are driven by renewable sources of energy" [34].

Scatec ASA is "a renewable energy solutions provider, developing, building, investing in, and operating renewable energy plants across different continents" that "derives maximum revenue from South Africa, followed by Botswana, Tunisia, Netherlands" [168]. The company maintains "4.6 GW in operation and under construction across four continents" with "close to 800 passionate employees" [34].

Solaria Energia y Medio Ambiente SA is "a solar PV power generation company" that "owns, manages and operates PV plants in Spain, Italy, Uruguay, Greece and Brazil" [169]. The company "manages plants in Spain, Italy, Portugal, Uruguay and Greece and has a pipeline of more than 14,200 MW" with target to "have installed 18 GW of emission-free energy by 2030" [34].

Terna Energy S.A is "a Greece based company mainly engaged in the growing renewable energy" that "promotes the development and implementation of solar, wind, geothermal, battery storage and green hydrogen technologies" [170]. The utility had "installed capacity of 905.23MW" by 2022 across "wind parks, hydroelectric projects, pumped storage projects, hybrid stations and photovoltaics" in "Greece, Bulgaria and Poland" [34].

Ørsted is a "Danish company" that was "named Dong Energy until the sale of all its oil and gas fields to Ineos in 2017" and "operated 9.9 gigawatts of offshore wind farms at the end of 2024" [165]. The company "develops, constructs, and operates offshore and onshore wind farms, solar farms, energy storage facilities" with approximately "8,000 people" employed and plans to "reach 50GW installed renewable capacity" by 2030 [34].

Table A.1 shows key characteristics of the final sample, revealing the diversity in scale, renewable portfolio composition, and business models across the 14 utilities. Companies range from specialised renewable energy developers to large vertically integrated utilities, with installed capacities spanning from under 1 GW to over 100 GW and renewable portfolios from 26% to 100%.

Table A.1: Sample Company Characteristics

Company	Country	Installed Capacity (MW)	Renewable Portfolio (%)*	Employee Count	Business Model Type
Akenerji	Turkey	1,224 (CDP)	26% (CDP)	Not specified	Generation-Focused
Arendals Fossekompani	Norway	Not specified	Not specified	Not specified	Investment Company
Atlantica	United Kingdom	2,464 (CDP)	86% (CDP)	Not specified	Infrastructure Investment
CEZ	Czech Republic	Not specified	Not specified	28,000+ (CDP)	Vertically Integrated Utility
EDF	France	116,900 (CDP)	90% (CDP)	171,490 (CDP)	Vertically Integrated Utility
EDP	Portugal	26,000 (CDP)	79% (CDP)	13,200 (CDP)	Vertically Integrated Utility
ERG	Italy	2,944 (CDP)	100% (CDP)	Not specified	Renewable Energy Specialist
Endesa	Spain	22,044 (CDP)	42% (CDP)	9,258 (CDP)	Vertically Integrated Utility
PGE	Poland	Not specified	Not specified	40,000 (CDP)	Vertically Integrated Utility
Romande Energie	Switzerland	Not specified	100% (CDP)	1,250+ (CDP)	Vertically Integrated Utility
Scatec	Norway	4,600 (CDP)	100% (CDP)	800 (CDP)	Renewable Energy Specialist
Solaria	Spain	Not specified	100% (Morningstar)	Not specified	Renewable Energy Specialist
Terna Energy	Greece	905 (CDP)	100% (CDP)	Not specified	Renewable Energy Specialist
Ørsted	Denmark	14,700 (Morningstar)	~99% (CDP)	8,000 (CDP)	Renewable Energy Specialist

*** Percentage of renewable energy within electricity portfolio**

Business model categories reflect distinct operational approaches: vertically integrated utilities handle the full electricity value chain from generation through distribution to retail; renewable energy specialists focus primarily on developing and operating clean energy assets; investment companies hold portfolios of energy infrastructure assets; generation-focused companies concentrate on power production activities. Infrastructure investment companies develop and own sustainable infrastructure projects across multiple regions.

B Sustainability Report Sources

This appendix provides direct links to all sustainability reports analysed in this research. The 2021-2022 period is before CSRD implementation, resulting in varied report titles and structures across companies. Some organisations published integrated reports, others focused ESG performance reports, and several companies used report different between years. This variation shows the pre-standardised reporting landscape that characterises the study period. See Table B.1 below.

Table B.1: Sustainability Report URLs by Company and Year

Company	2021 Report	2022 Report
Akenerji	Integrated Report	Integrated Report
Arendals	Annual Report	Annual Integrated Report
Atlantica	ESG Supplement to Annual Report	Integrated Annual Report
CEZ	Sustainability Report	Sustainability Report
EDF	CSR Reports and Indicators	Sustainability Report
EDP	Sustainability Report	Integrated Report
ERG	Consolidated Non-Financial Statement Report	Consolidated Non-Financial Statement Report
Endesa	Non-Financial Information and Sustainability Statement	Non-Financial Information and Sustainability Statement
PGE	Integrated Report	Integrated Report
Romande	Sustainability Report	Sustainability Report
Scatec	ESG Performance Report	ESG Performance Report
Solaria	ESG Report	Sustainability Report
Terna	CSR Report	CSR Report
Ørsted	Sustainability Report	Sustainability Report

C Performance Data and Scores

C.1 Required Data

This section presents the raw performance data extracted from CDP and RE sources used in the environmental performance component of the GRAT. All data comes from RE except for the CDP emission intensity values, which provide the self-reported comparison component for transparency assessment. Revenue figures are expressed in million US dollars, while Scope 1 and 2 emissions are measured in tons CO2 equivalent.

Missing data points are marked with asterisks (*) and indicate cases where target years or reduction percentages were unavailable for specific years. When this occurred, values from adjacent years were used to maintain data completeness for the performance calculations.

The industry benchmarks used for calculating relative performance scores were derived from the 27-company European utility dataset. These sector averages provide the reference points for emission intensity and renewable energy scoring:

- Average Emission Intensity 2021: 1,265.07 tCO2e/million USD
- Average Emission Intensity 2022: 686.67 tCO2e/million USD
- Average Renewable Intensity 2021: 6,801.5 units/million USD
- Average Renewable Intensity 2022: 6,229.29 units/million USD

Table C.1: Complete Performance Dataset

Company	CDP Int.21	CDP Int.22	Eikon Int.20	Eikon Int.21	Eikon Int.22	Ren. 20	Ren. 21	Ren. 22	Target Yr 20	Target Yr 21	Target Yr 22	Target % 20	Target % 21	Target % 22
Akenerji	3.5E3	1.1E3	4.7E3	5.3E3	1.3E3	1.1E4	9.3E3	3.0E3	2035	2030	2030	50	55	50
Arendals	5.8E-1	3.3E-1	2.5E0	9.2E0	8.7E0	5.8E1	–	–	2050	2030	2030	100	98	42
Atlantica	1.7E3	1.9E3	1.9E3	1.7E3	1.9E3	1.0E3	8.8E2	1.5E3	2035	2035	2035	70	70	70
CEZ	1.9E3	1.5E3	2.4E3	1.9E3	1.5E3	2.0E3	1.5E3	9.5E2	2025	2030	2025*	33	50	50*
EDF	2.8E2	1.6E2	3.3E2	2.9E2	1.6E2	3.6E3	1.3E0	8.4E-1	2030	2023	2023	50	23	23
EDP	6.0E2	4.5E2	6.5E2	6.2E2	4.5E2	1.1E4	9.6E3	7.4E3	2030	2025	2025	90	70	70
ERG	6.4E2	7.9E0	1.1E3	1.4E3	7.8E0	–	2.2E4	2.2E4	2050	2030	2025	55	55	45
Endesa	2.2E3	4.0E2	5.2E2	4.7E2	4.0E2	2.3E3	1.9E3	1.2E3	2023	2023	2023*	70	75	75*
PGE	3.7E6	4.4E3	4.8E3	5.6E3	4.3E3	7.3E2	8.3E2	6.9E2	2050	2030	2030	100	80	80
Romande	2.0E-2	2.2E4	1.0E1	1.3E1	1.3E1	–	–	–	2023*	2023*	2023	5*	5*	5
Scatec	1.1E1	1.9E1	3.1E1	2.9E1	3.1E1	6.4E1	2.6E1	1.4E2	2030	2030	2030	50	50	95
Solaria	9.0E0	1.1E1	4.0E0	9.3E0	1.1E1	2.8E4	2.9E4	3.4E4	2021	2030	2030	14	100	100
Terna	8.9E0	9.7E0	5.2E2	6.0E2	5.8E2	8.3E1	8.5E1	6.7E1	2030*	2030*	2030	28*	28*	46
Ørsted	1.8E2	1.4E2	2.4E2	1.9E2	1.6E2	1.3E4	1.0E4	8.4E3	2023	2023	2025	20	20	97.83

Note: CDP Int. = CDP Emission Intensity (tCO2e/million USD), Eikon Int. = Eikon Emission Intensity (tCO2e/million USD), Ren/Rev = Renewable Energy per Revenue (units/million USD), Target Yr = Emission Reduction Target Year, Target % = Emission Reduction Target Percentage. (--) indicates unavailable data.

C.2 Component Scores

Table C.2 below presents individual component scores for all companies across both years, calculated before ensemble integration. Components follow a 0-100 scoring scale. The Suspicious Changes Penalty can subtract up to 10 points, while the Renewable Energy Bonus adds up to 10 points. Companies are scored against industry benchmarks and theoretical thresholds rather than being normalised within the sample, following the methodology from Chapter 3.3.1.

Table C.2: Individual Performance Component Scores

Company	Year	Emission Intensity	Goal Achiev.	Target Amb.	Transp.	Sus. Changes Penalty	Ren. Energy Bonus
Akenerji	2021	0	0	61.11	0	0	4.87
Akenerji	2022	5.43	100	62.5	40	-2.73	0
Arendals	2021	100	0	100	0	-2.16	0
Arendals	2022	100	61.46	52.5	0	-4	0
Atlantica	2021	25.81	100	50	99.86	0	0
Atlantica	2022	0	0	53.85	95.62	0	1.76
CEZ	2021	19.6	100	55.56	100	-4	0
CEZ	2022	0	86.44	100	100	-10	0
EDF	2021	100	86.26	100	71.38	-4	0
EDF	2022	100	100	100	100	0	0
EDP	2021	100	15.44	100	71.79	-3.78	5.04
EDP	2022	68.31	80.79	100	100	0	4.17
ERG	2021	38.33	0	61.11	0	0	8
ERG	2022	100	100	100	100	-3.45	8
Endesa	2021	100	16.72	100	100	0	0
Endesa	2022	76.61	20.75	100	100	-10	0
PGE	2021	0	0	88.89	100	-3.6	1.13
PGE	2022	0	100	100	100	0	0
Romande	2021	100	0	25	0	-10	0
Romande	2022	100	0	50	100	0	0
Scatec	2021	100	59.48	55.56	0	0	0
Scatec	2022	100	0	100	0	0	2
Solaria	2021	100	0	100	71.93	-10	9.02
Solaria	2022	100	0	100	100	0	9.18
Terna	2021	100	0	31.11	0	0	1.02
Terna	2022	51.78	27.28	57.5	0	0	0
Ørsted	2021	100	100	100	70.91	0	5.28
Ørsted	2022	100	33.17	100	36.37	-2.8	4.76

Note: Goal Achiev. = Goal Achievement, Target Amb. = Target Ambition, Transp. = Transparency, Sus. Changes = Suspicious Changes, Ren. Energy = Renewable Energy.

C.3 Ensemble statistics

Table C.3 below shows the ensemble statistics calculated across all 2,285 valid weight combinations for each company-year observation. The ensemble method tests different weight allocations within theoretical constraints, creating a distribution of scores for each case. The median scores become the final performance values used in the greenwashing analysis. Standard deviation and quartile ranges show how much scores change when weights vary.

Table C.3: Performance Ensemble Statistics

Company	Year	Mean	Median	Std Dev	Min	Max	Q25	Q75	IQR
Akenerji	2021	14.8	14.98	2.59	7.76	21.09	13.31	16.65	3.34
Akenerji	2022	40.96	41.37	3.28	31.46	47.1	38.58	43.71	5.13
Arendals	2021	52.43	52.58	3.93	44.4	63.49	49.85	55.31	5.46
Arendals	2022	59.28	59.78	3.77	46.89	66.47	56.86	62.16	5.3
Atlantica	2021	54.65	54.82	2.79	47.18	61.76	52.59	56.68	4.09
Atlantica	2022	19.85	19.7	4.59	8.88	33.34	16.66	22.96	6.3
CEZ	2021	49.64	49.78	3.08	41.84	57.84	47.34	51.9	4.56
CEZ	2022	40.74	40.43	4.85	31.56	54.07	36.79	44.36	7.57
EDF	2021	80.84	80.91	0.9	78.04	83.1	80.27	81.47	1.2
EDF	2022	90.91	90.91	0	90.91	90.91	90.91	90.91	0
EDP	2021	66.54	66.68	3.2	56.95	74.37	64.38	68.73	4.35
EDP	2022	77.66	77.59	1.68	73.49	82.38	76.44	78.85	2.41
Endesa	2021	68.43	68.95	3.56	57.6	77.28	65.92	71.22	5.3
Endesa	2022	51.68	52.16	3.48	40.55	58.43	49.46	54.32	4.86
ERG	2021	31.99	32.04	2.02	26.29	36.92	30.58	33.5	2.92
ERG	2022	95.04	95.04	0	95.04	95.04	95.04	95.04	0
Ørsted	2021	92.94	93.07	1.06	89.63	94.39	92.27	93.86	1.59
Ørsted	2022	68.6	68.59	2.59	63.07	75.83	66.71	70.48	3.77
PGE	2021	22.37	22.51	5.82	7.15	38.06	18.47	26.34	7.87
PGE	2022	53.49	53.64	4.86	45.45	66.36	50	57.27	7.27
Romande	2021	32.57	32.5	4.27	20.91	41.36	29.32	35.91	6.59
Romande	2022	55.43	55	3.78	48.18	66.82	52.73	58.18	5.45
Scatec	2021	62.9	63.43	3.67	50.72	69.57	60.56	65.72	5.16
Scatec	2022	56.22	56.36	3.93	48.18	67.27	53.64	59.09	5.45
Solaria	2021	60.36	60.48	3.84	48.75	69.83	57.84	63.07	5.23
Solaria	2022	72.26	72.89	4.27	59.25	82.89	69.25	75.62	6.37
Terna	2021	43.63	43.57	4.15	32.26	52.61	40.56	46.93	6.37
Terna	2022	36.5	36.52	1.65	31.71	40.74	35.38	37.68	2.3

Standard deviations range from perfect agreement (0.0 for top performers like EDF 2022 and ERG 2022) to higher uncertainty (5.82 for PGE 2021). This suggests stronger reliability for companies with clear performance profiles and more sensitivity for transitional cases. Q25 and Q75 represent the 25th and 75th percentiles of the score distribution, while IQR measures the spread between these quartiles. These statistics reveal how stable scores remain across different weighting scenarios.

D NLP Modules and Verification

The sections below explain how each module works conceptually and show verification examples of their accuracy in detecting relevant patterns. The verifications (in grey, see below) are verification print statement copied directly from the modules. For the actual code implementation, module specifications, and technical documentation, see the GitHub repository referenced in Appendix K.

D.1 Green Terms Module

This module counts environmental words in sustainability reports to measure how much companies talk about green topics.

The word lists include two main types:

- **Natural (inherently) green terms**: Words that are naturally environmental. Examples are 'renewable', 'sustainable', 'biodiversity', 'solar power', 'carbon neutral'.
- **Context-dependent terms**: Neutral words like 'emissions', 'CO2', or 'carbon footprint' that only count as green when paired with the right context words: '**reduced** emissions', '**captured** CO2', '**improved** footprint'.

The extraction works in stages to avoid double-counting. Longer term/phrases get picked up first ('renewable energy generation'), then shorter ones ('renewable'). This stops the same text/terms from being counted twice.

Negation filtering catches green terms used in negative ways. Examples:

- '**Lack of** renewable energy investments' does not count as green.
- '**Unable to** meet sustainability targets' 'sustainability' is excluded.

The system checks different parts of sentences around each term to find various negation patterns: direct negations such as '**not** sustainable', failed actions such as 'could**n't** improve efficiency', and broader negative phrases.

The final calculation as follows:

- Green term frequency = (green words ÷ total content words) × 100.
- Total content words exclude stopwords, punctuation, and whitespace.
- Vocabulary diversity tracks unique environmental words found.
- Quality checks show how many terms got filtered out.

Outputs for Communication Scoring

- Green Communication Intensity dimension: This frequency percentage and vocabulary diversity is used in the main communication scoring system (PCG).

It measures environmental content coverage across climate, energy, and sustainability topics [89]. Providing a clear measure of how frequently environmental terms appear relative to the rest of the report.

D.1.1 Green Term Verification

The module accurately identifies a range of term types, from basic nouns such as "sustainability" to more complex phrases like "low carbon economy." It also handles context-dependent terms, correctly classifying neutral words like "waste" when they appear in green contexts. See below:

VALID TERMS IN TEXT ORDER:

————————————————————————-

1. NOUN TERM: sustainability

*Context: the scope of our announced **sustainability** goals , we focused on*

2. NOUN TERM: adaptation

*Context: determined our focus areas as **adaptation** to climate change and transition*

2. VERB TERM: transition

*Context: adaptation to climate change and **transition** to a low carbon economy*

2. MULTIWORD ADJECTIVE TERM: low carbon

*Context: change and transition to a **low carbon** economy , zero waste and*

5. CONTEXT-DEPENDENT NOUN (NEGATIVE): zero waste

(Context: 'zero' -> Neutral: 'waste')

*Context: low carbon economy , zero **waste** and circular economy , combating*

The negation filter excludes terms that should not count as green communication by applying methods such as direct negation and dependency parsing. By doing this, green terms that are used in a negative context are not counted since this is not green communication. However, it occasionally filters out valid terms due to the complexity of sentences and cannot detect double negatives (see: no large environmental spills).

EXAMPLES OF NEGATED TERMS (EXCLUDED FROM COUNT):

————————————————————————-

1. ADJECTIVE TERM: environmental

Negated by: No (type: direct_negation_head_subtree in head_subtree scope)

*Context: No large **environmental** spills occurred in terms of*

2. NOUN TERM: improvement

Negated by: not (type: spacy_neg_head_subtree in head_subtree scope)

*Context: Standards , amendments and **improvements** issued but not yet effective*

3. VERB TERM: transition

Negated by: without (type: direct_negation_ancestors in ancestors scope)

*Context: Changes without any **transition** requirement , material changes in*

4. ADJECTIVE TERM: recoverable

Negated by: not (type: spacy_neg_head_subtree in head_subtree scope)

*Context: carrying amount may not be **recoverable** .*

D.2 Green Term Context Classification Module

This module classifies the context of already-identified green terms in the previous module across three dimensions to assess communication quality. It is added after the existing code for the Green Term Module.

The module applies three separate classification checks:

1. Temporal Scope Classification

- **Past context**: Terms describing completed actions: '**achieved** carbon neutrality in 2020', '**reduced** emissions by 15% last year'.
- **Present context**: Terms describing current/present status: 'currently **operating** 500MW of solar capacity', 'our renewable energy portfolio **generates**'.
- **Future context**: Terms describing planned actions: '**will** install wind turbines by 2030', 'targeting net zero emissions **by 2050**'.

The system finds governing (connected) verbs for each found green term using dependency parsing. It then analyses verb tense and searches for temporal markers within five tokens of the term. Temporal markers are adjusted for report year. 2021 reports treat 2021 as present, ≥2022 as future etc.

2. Quantification Classification The module detects two types of quantification connections:

- Direct quantifications: Specific numbers with units: **'2,284 MWh** of clean energy', '**€50 million** investment in renewables', '**25%** reduction in emissions'.
- Relative quantifications: Comparative terms without direct quantifications: '**doubled** renewable capacity', '**increased** efficiency'.

The code reconstructs quantifications from fragmented tokens: The code searches for numbers near currency symbols, units, or percentage signs within a few tokens, then combines them back into complete quantifications. It assesses syntactic connections using three methods: direct dependency relationships, proximity within sentences, and pattern recognition. The connection strength determines final classification:

- Highly quantified: Strong connection to currency, percentage, or unit quantifications.
- Partially quantified: Connection to relative quantifiers or weaker direct quantifications.
- Non-quantified: No meaningful/close proximity quantification connections found.

3. Evidence vs Aspirational Classification This classification identifies the difference between concrete evidence and future commitments:

- Strong evidence terms: Completed, verifiable actions: '**installed** 200MW solar capacity', '**achieved** ISO 14001 certification', '**measured** 15% emission reduction'.
- Moderate evidence terms: Ongoing or implemented actions: '**operates** renewable facilities', '**generates** clean energy', '**maintains** environmental standards'.
- Strong aspirational terms: Uncertain or conditional future plans: '**should** achieve neutrality', '**might** invest in renewables', '**vision** includes clean energy'.
- Moderate aspirational terms: Committed future intentions: '**will** install wind turbines', '**plan to** reduce emissions', '**strategy** targets carbon neutrality'.

The system uses targeted marker detection around semantic governors (main verb that controls a green term in a sentence). It avoids sentence-wide searches. Evidence markers include completed actions ('achieved', 'implemented'), verified outcomes ('certified', 'audited'), and measured results ('demonstrated', 'recorded'). Aspirational markers include modal uncertainty ('should', 'could'), future commitments ('will', 'plan to'), and strategic language ('strategy', 'roadmap').

Outputs for Communication Scoring

- Temporal Orientation dimension: Past/present/future percentages (for each found green term) measure whether companies focus on achievements versus promises.
- Substantiation Weakness dimension: Quantification intensity scores assess how well environmental claims are supported with specific numbers. Also, Evidence vs aspirational intensity scores distinguish between concrete proof and future commitments.

The module generates intensity scores that calculate weighted percentages. Strong classifications get 1.5 points; moderate classifications get 1.0 point. Formula: ((strong count $\times$ 1.5) + (moderate count $\times$ 1)) $\div$ total green terms $\times$ 100. For example, if a report has 10 strong evidence terms, 5 moderate evidence terms, and 100 total terms, the evidence intensity score equals (10 $\times$ 1.5 + 5 $\times$ 1.0) $\div$ 100 $\times$ 100 = 20%. Same for quantification but then 1.5 point for highly quantified and 1.0 point for partially quantified. These scores enable comparison across documents.

D.2.1 Context Classification Verification

D.2.1.1 Temporal

This module consists of three classification parts that analyse green terms in different ways. The temporal analysis shows how the system sorts terms by temporal scope: past actions like "determined focus areas," present work like ongoing renewable projects, and future plans like upcoming efficiency projects.

PAST EXAMPLES:

——————————————————————————-

1. Term: 'high efficiency'

Document: Akenerji_Elektrik_ Uretim_ AS_ 2021 (Report Year: 2021)

*Context: Our natural gas combined cycle, wind and 7 hydroelectric power plants are run with principles of 'high profitability and **high efficiency**' and has been adding power to our country for 33 years...*

2. Term: 'low carbon'

Document: Akenerji_Elektrik_ Uretim_ AS_ 2021 (Report Year: 2021)

*Context: We determined our focus areas as adaptation to climate change and transition to a **low carbon** economy, zero waste and circular economy, combating water scarcity, gender equality, cooperation in order to develop human resource competencies, reducing the carbon and water footprint of suppliers, energy efficiency, generation efficiency by reducing risks and sustainable management by augmenting innovation.*

3. Term: 'circular economy'

Document: Akenerji_Elektrik_ Uretim_ AS_ 2021 (Report Year: 2021)

*Context: We determined our focus areas as adaptation to climate change and transition to a low carbon economy, zero waste and **circular economy**, combating water scarcity, gender equality, cooperation in order to develop human resource competencies, reducing the carbon and water footprint of suppliers, energy efficiency, generation efficiency by reducing risks and sustainable management by augmenting innovation.*

PRESENT EXAMPLES:

——————————————————————————-

1. Term: 'energy efficiency'

Document: Akenerji_Elektrik_ Uretim_ AS_ 2021 (Report Year: 2021)

*Context: We determined our focus areas as adaptation to climate change and transition to a low carbon economy, zero waste and circular economy, combating water scarcity, gender equality, cooperation in order to develop human resource competencies, reducing the carbon and water footprint of suppliers, **energy efficiency**, generation efficiency by reducing risks and sustainable management by augmenting innovation.*

2. Term: 'renewable energy'

Document: Akenerji_Elektrik_ Uretim_ AS_ 2021 (Report Year: 2021)

*Context: Akenerji has become the representative of foresight and stability in the sector with its breakthroughs in **renewable energy**, which is gaining more and more importance in the world and in our country.*

3. Term: 'solar energy'

Document: Akenerji_Elektrik_ Uretim_ AS_ 2021 (Report Year: 2021)

*Context: Akenerji continues to conduct market research on projects with high capacity utilization and profitability for wind and **solar energy** to include in its portfolio.*

FUTURE EXAMPLES:

——————————————————————————-

1. Term: 'energy efficiency'

Document: Akenerji_Elektrik_ Uretim_ AS_ 2021 (Report Year: 2021)

*Context: We are putting our signature to leading projects in **energy efficiency**.*

2. Term: 'renewable energy'

Document: Akenerji_Elektrik_ Uretim_ AS_ 2021 (Report Year: 2021)

*Context: The company continues its efforts to reshape its generation strategies and to take necessary actions to create maximum benefit from **renewable energy** sources.*

3. Term: 'energy efficiency'

Document: Akenerji_Elektrik_ Uretim_ AS_ 2021 (Report Year: 2021)

*Context: The Company focuses on producing new ideas and projects that will enhance **energy efficiency** for industrial and commercial customers, with energy systems optimization and management services.*

D.2.1.2 Quantification

The quantification examples show three levels of quantified support. Highly quantified terms connect to specific figures like "570,000 tons," partially quantified ones use relative words like "more efficient," while non-quantified terms lack supporting quantification entirely.

HIGHLY QUANTIFIED EXAMPLES:

———————————————————————————-

1. 'emission reduction' (multiword_noun)

*Context: 550,000 trees to the nature with the 570,000 tons of **emission** **reduction** certificates it sold in 2020 .*

Confidence: 1.000

Quantifications: '550,000 trees' (meaningful_count, pattern, strength=0.90), '570,000 tons' (unit, pattern, strength=0.90)

Scores: Highly=3.60, Partially=0.00

2. 'emission reduction' (multiword_noun)

*Context: these announced commitments cover less than 20 % of the **emission** **reductions** required by 2030 to make the targeted 1.5 degree temperature*

Confidence: 1.000

Quantifications: 'less' (relative, pattern, strength=0.90), '20 %' (percentage, pattern, strength=0.90)

Scores: Highly=1.80, Partially=0.90

3. 'net zero' (multiword_noun)

*Context: the targeted 1.5 degree temperature rise achievable and to reach **net** **zero** emissions by 2050 .*

Confidence: 1.000

Quantifications: 'less' (relative, pattern, strength=0.90), '20 %' (percentage, pattern, strength=0.90), 'zero emissions' (meaningful_count, syntactic, strength=0.80)

Scores: Highly=3.40, Partially=0.90

PARTIALLY QUANTIFIED EXAMPLES:

———————————————————————————-

1. 'efficient energy' (multiword_adjective)

*Context: We continue to take innovative steps for more **efficient** **energy** generation .*

Confidence: 0.450

Quantifications: 'more' (relative, syntactic, strength=0.90)

Scores: Highly=0.00, Partially=0.90

2. 'positive energy' (context_dependent_noun)

*Context: mission is to provide safe , reliable , and positive **energy** to its customers and society as a whole ; its*

Confidence: 0.450

Quantifications: 'positive' (relative, syntactic, strength=0.90)

Scores: Highly=0.00, Partially=0.90

3. 'renewable energy' (multiword_noun)

*Context: foresight and stability in the sector with its breakthroughs in **renewable** **energy** , which is gaining more* *and more importance in the*

Confidence: 0.400

Quantifications: 'gaining' (relative, syntactic, strength=0.80)

Scores: Highly=0.00, Partially=0.80

NON QUANTIFIED EXAMPLES:

——————————————————————————-

1. 'high efficiency' (multiword_ adjective)

*Context: plants are run with principles of ' high profitability and **high** **efficiency** ' and has been adding power to our country for*

Confidence: 0.000

Quantifications: None

2. 'energy efficiency' (multiword_ adjective)

*Context: We are putting our signature to leading projects in **energy** **efficiency** .*

Confidence: 0.000

Quantifications: None

3. 'energy efficiency' (multiword_ adjective)

*Context: , reducing the carbon and water footprint of suppliers , **energy** **efficiency** , generation efficiency by reducing risks and sustainable management by*

Confidence: 0.000

Quantifications: None

D.2.1.3 Evidence Aspirational

The evidence classification separates concrete proof from aspirational claims/promises. Strong evidence includes completed actions like "produced 5,120,168MWh," while aspirational terms use uncertain language like "will enhance" or appear in project contexts without clear outcomes.

STRONG EVIDENCE EXAMPLES:

——————————————————————————-

1. 'environmentally friendly' (direct)

*Context: With our balanced, sustainable, efficient and **environmentally friendly** portfolio, we produced 5,120,168MWh of gross electricity in 2021.*

Confidence: 0.900

Semantic Governor: 'produced' (VERB)

Evidence markers: 'produced' (strong_ evidence, governor_ modifiers, direct_ modification_ semantic_ governor)

2. 'environmental management' (direct)

*Context: **Environmental Management** Systems recertification audits for the Head Office, Ayyıldız WPP, Uluabat HEPP, Burç HEPP, Bulam HEPP, Feke I HEPP, Feke II HEPP, Himmetli HEPP, Gökkaya HEPP and Erzin NGCC plant were successfully completed and documents were renewed.*

Confidence: 0.900

Semantic Governor: 'completed' (VERB)

Evidence markers: 'completed' (strong_evidence, governor_modifiers, direct_modification_semantic_governor)

... and 37 more strong evidence terms

MODERATE EVIDENCE EXAMPLES:

——————————————————————————-

1. 'wind power' (direct)

*Context: Akenerji also evaluates the capacity increase opportunities in existing plants, The investment process was initiated in 2016 in order to boost the installed power in Ayyıldız **Wind Power** Plant to 28.2 MW from 15 MW.*

Confidence: 0.600

Semantic Governor: 'boost' (VERB)

Evidence markers: 'installed' (strong_evidence, governor_modifiers, proximity_2_semantic_governor)

2. 'wind power' (direct)

*Context: Ayyıldız **Wind Power** Plant with an installed capacity of 15 MW was activated in September.*

Confidence: 0.450

Semantic Governor: 'activated' (VERB)

Evidence markers: 'activated' (moderate_evidence, governor_modifiers, direct_modification_semantic_governor)

... and 49 more moderate evidence terms

NEUTRAL EXAMPLES:

——————————————————————————-

1. 'high efficiency' (direct)

*Context: Our natural gas combined cycle, wind and 7 hydroelectric power plants are run with principles of 'high profitability and **high efficiency**' and has been adding power to our country for 33 years...*

Confidence: 0.000

Semantic Governor: 'run' (VERB)

Markers: None (classified as neutral)

2. 'low carbon' (direct)

*Context: We determined our focus areas as adaptation to climate change and transition to a **low carbon** economy, zero waste and circular economy, combating water scarcity, gender equality, cooperation in order to develop human resource competencies, reducing the carbon and water footprint of suppliers, energy efficiency, generation efficiency by reducing risks and sustainable management by augmenting innovation.*

Confidence: 0.000

Semantic Governor: 'determined' (VERB)

Markers: None (classified as neutral)

... and 487 more neutral terms

MODERATE ASPIRATIONAL EXAMPLES:

——————————————————————————-

1. 'energy efficiency' (direct)

*Context: We are putting our signature to leading projects in **energy efficiency**.*

Confidence: 0.300

Semantic Governor: 'putting' (VERB)

Aspirational markers: 'project' (moderate_aspirational, token_modifiers, proximity_2_main_token)

2. 'carbon management' (direct)

*Context: The **Carbon Management** Project was launched.*

Confidence: 1.000

Semantic Governor: 'launched' (VERB)

Evidence markers: 'launched' (strong_evidence, governor_modifiers, direct_modification_semantic_governor)

Aspirational markers: 'project' (moderate_aspirational, token_modifiers, proximity_2_main_token), 'project' (moderate_aspirational, token_modifiers, direct_modification_main_token), 'project' (moderate_aspirational, governor_modifiers, direct_modification_semantic_governor)

... and 56 more moderate aspirational terms

STRONG ASPIRATIONAL EXAMPLES:

——————————————————————————-

1. 'energy efficiency' (direct)

*Context: The Company focuses on producing new ideas and projects that will enhance **energy efficiency** for industrial and commercial customers, with energy systems optimization and management services.*

Confidence: 1.000

Semantic Governor: 'enhance' (VERB)

Aspirational markers: 'will' (moderate_aspirational, token_modifiers, proximity_2_main_token), 'will' (moderate_aspirational, token_modifiers, sibling_of_main_token), 'will' (moderate_aspirational, governor_modifiers, direct_modification_semantic_governor)

2. 'efficiency project' (direct)

*Context: Energy Services, together with **efficiency projects** and engineering services, has proven once more that the Akenerji family is a competent leader in every facet of the energy sector.*

Confidence: 0.800

Semantic Governor: 'proven' (VERB)

Aspirational markers: 'project' (moderate_aspirational, token_modifiers, direct_modification_main_token), 'project' (moderate_aspirational, token_modifiers, sibling_of_main_token)

... and 45 more strong aspirational terms

D.3 Sentiment Analysis Module

This module measures environmental messaging sentiment using Climate-BERT, a specialised model trained on climate texts. Climate-BERT classifies sustainability content into three categories: opportunity (positive climate sentiment), risk (negative climate sentiment), and neutral (other content). The process works in two steps. First, it finds all sentences with green terms from the previous module. Then Climate-BERT scores each sentence. The final calculation uses:

Sentiment Score = Opportunity Score - Risk Score

Values range from -1 (fully risk-focused) to +1 (fully opportunity-focused).

The system extracts sentences containing green terms and analyses sentiment at two levels. Document-level sentiment combines all sustainability sentences per report, producing average sentiment scores. Topic-specific weighted sentiment calculates separate scores for renewable energy and climate emissions categories. When sentences contain multiple topic terms, sentiment contributions get weighted by term frequency.

Analysis targets renewable energy and climate emissions because these topics carry distinctly positive sentiment that shapes public perception [135]. Research shows that "receptivity to green communication was the most important factor affecting green attitudes" [248]. The focus remains 'on sustainability messages' and environmental content coverage, examining topics like climate, energy, and sustainability [89].

Renewable energy terms include examples as 'solar power', 'wind turbines', 'clean energy generation'. Climate emissions terms cover terms such as 'carbon neutral', 'emission reduction', 'decarbonization'.

Outputs for Communication Scoring

- Green Communication Intensity dimension: Overall sentiment scores measure whether companies frame environmental topics positively or negatively. Also, weighted sentiment scores for renewable energy and climate emissions assess presentation of areas that drive public environmental attitudes.

D.3.1 Sentiment Verification

This module uses ClimateBERT that scores sentences containing green terms. The examples show how identical terms like "renewable energy" get different sentiment scores based on context: positive when discussing growth opportunities, negative when mentioning market risks or regulatory challenges.

OPPORTUNITY EXAMPLES (201 total):

——————————————————————-

1. Term: 'alternative energy' | Topic: renewable_energy

Sentiment Score: +0.958 | Confidence: 0.962

Text: As a company that focuses on power plant investments based on renewable resources in order to create resource diversity and cost advantage, we have ra...

2. Term: 'alternative energy' | Topic: renewable_energy

Sentiment Score: +0.961 | Confidence: 0.964

Text: Within the framework of the actions taken on energy efficiency and alternative energy resources, Akenerji Energy Services, thanks to the special solut...

3. Term: 'clean energy' | Topic: renewable_energy

Sentiment Score: +0.743 | Confidence: 0.749

Text: The 'VISION 2030 - Clean Energy for Tomorrow' strategy is focused on the dynamic transformation of the generation portfolio to a low-emission portfoli...

RISK EXAMPLES (82 total):

——————————————————————-

1. Term: 'renewable energy' | Topic: renewable_energy

Sentiment Score: -0.546 | Confidence: 0.565

Text: The Renewable Energy Resources Support Mechanism (RERSM) continued to maintain its importance for independent producers due to the fluctuations in ele...

2. Term: 'renewable energy' | Topic: renewable_ energy

Sentiment Score: -0.974 | Confidence: 0.976

Text: The Management prioritizes reports, and monitors the risks in line with the Risk Appetite

D.4 Vague and Hedge Words Module

This module detects unclear language in sustainability reports to measure communication precision. Companies use hedge words to express uncertainty and vague words to avoid specificity.

The word lists include two main types:

- Hedge words: Express tentativeness. Strong hedges like 'might', 'could', 'appears', 'potentially'. Mild hedges like 'often', 'generally', 'somewhat', 'typically'.
- Vague words: Lack precision across categories. Temporal vagueness like 'soon', 'eventually'. Scope ambiguity like 'various', 'several'. Commitment vagueness like 'working towards', 'aiming to'. Degree vagueness like 'significant', 'substantial'.

Context-dependent analysis determines when words become specific:

- 'Improved performance' counts as vague.
- 'Improved by 25%' or 'improved compared to 2021' do not count due to quantification or comparison.

The extraction works in stages to avoid double-counting. Multi-word phrases get picked up first, then single words. Context-dependent words go through additional checks for surrounding numbers, percentages, or comparative phrases with close proximity to the word.

The final calculation works like this:

- Hedge density = (hedge words ÷ total content words) × 100.
- Vague density = (vague words ÷ total content words) × 100.
- Intensity scores weight strong unclear language 1.5x more than mild (similar as in context classification).

Outputs for Communication Scoring

- Language Precision dimension: Hedge densities measure careful word choice that avoids overconfident claims while vague densities assess imprecision.

This measures communication precision, showing whether companies choose words carefully or use empty language without supporting evidence.

D.4.1 Vague and Hedge Verification

The module tracks different types of unclear language: vague words that lack precision or timeline specificity, and hedge words that express uncertainty. This examines if companies use empty language or appropriate caution in their environmental claims.

D.4.1.1 Vague Words

The system catches both strong vague terms like "significant" and "various" that mean nothing specific, but also mild vague words like "efforts" and "reducing" that lack details.

STRONG VAGUE WORDS (211 total):

——————————————————————————-

1. 'significant' (lemma: significant)

Context: Reinvigorating economic activities around the world led to a significant rise in energy demand and the revival of energy markets

2. 'strong' (lemma: strong)

Context: Globally , we have seen a strong growth trend in electricity demand in the manufacturing industry ,

3. 'some' (lemma: some)

Context: approximately 5.4 TWh at the end of 2021 , growing some 10 times compared to the previous year .

4. 'significant' (lemma: significant)

Context: Undoubtedly , we are proud of these significant achievements .

5. 'various' (lemma: various)

Context: Thanks to our collaborations with various university clubs , we have helped young talents realize their

MILD VAGUE WORDS (794 total):

——————————————————————————-

1. 'efforts' (lemma: effort)

Context: , 2021 has been recorded as a year in which efforts were made predominently to repair the damages caused by the

2. 'develop' (lemma: develop)

Context: water scarcity , gender equality , cooperation in order to develop human resource competencies , reducing the carbon and water footprint

3. 'reducing' (lemma: reduce)

Context: , cooperation in order to develop human resource competencies , reducing the carbon and water footprint of suppliers , energy efficiency

4. 'innovation' (lemma: innovation)

Context: generation efficiency by reducing risks and sustainable management by augmenting innovation .

5. 'valuable' (lemma: valuable)

Context: express my gratitude to all our stakeholders , especially our valuable employees , financiers and investors , who have always contributed

D.4.1.2 Vague commitments

Commitments are classified based on the presence or absence of clear timelines. Terms such as "evaluating" or "preparing" are marked as vague unless paired with time references like "in 2021."

VAGUE COMMITMENTS (WITHOUT TIMELINE) - 5 total:

——————————————————————————-

1. 'evaluating'

Context: reduce consumption and make improvements by determining energy consumption and evaluating performance .

2. 'preparing'

Context: 's border with the Czech Republic and owns a developer preparing the construction of wind parks.

COMMITMENTS (WITH TIMELINE) - 3 total:

——————————————————————————-

1. 'preparing'

Context: By preparing our first integrated report in 2021 , we raised our

2. 'establishing'

Context: utilizing domestic resources , reducing foreign dependency in energy and establishing largecapacity renewable energy plants , continued in 2021 as well

D.4.1.3 Hedge Words

Strong hedge words like "possible" and "may" show higher uncertainty, while mild hedging like "approximately" and "significantly" add caution without being too uncertain.

STRONG HEDGE WORDS (147 total):

——————————————————————————-

1. 'possible' (lemma: possible)

Context: ...to minimize the [possible] risks our activities...

2. 'provisional' (lemma: provisional)

Context: ...2020 in the [provisional] article 1 of...

3. 'may' (lemma: may)

Context: ...2021 from leases [may] not be taken...

4. 'possible' (lemma: possible)

Context: ...as much as [possible] and consult different...

5. 'may' (lemma: may)

Context: ...business areas that [may] be an opportunity...

MILD HEDGE WORDS (99 total):

——————————————————————————-

1. 'significantly' (lemma: significantly) [POS: ADV, DEP: advmod]

Context: ..., we have [significantly] improved the power...

2. 'accordingly' (lemma: accordingly)

Context: ...license was transferred [accordingly] .

3. 'approximately' (lemma: approximately) [POS: ADV, DEP: advmod]

Context: ...capacity to provide [approximately] 2.5% of...

4. 'regularly' (lemma: regularly)

Context: .... With its [regularly] rising international sales...

5. 'significantly' (lemma: significantly) [POS: ADV, DEP: advmod]

Context: ...Europe and contributes [significantly] to the development...

D.5 Document Similarity Module

This module compares sustainability reports across years to catch copy-pasting versus new authentic content. Companies that repeat the same text score high similarity while those spending more time writing new/different reports score lower.

The analysis uses three approaches:

- Term Frequency-Inverse Document Frequency (TF-IDF) similarity: Compares whole documents using word frequencies. Shows overall content overlap between years.
- Jaccard similarity: Measures shared vocabulary as (common words ÷ total unique words). Reveals whether companies recycle the same terms.
- Sentence matching: Pairs individual sentences between 2021 and 2022 reports using semantic similarity.

Sentence processing filters meaningful content first. Only sentences with more than 10 content words get analysed (excluding stopwords and punctuation). The system then matches each sentence from the shorter document with its best counterpart from the longer document using greedy algorithm.

The matching works step by step. Find the highest similarity score among all possible sentence pairs, select those pairs, record the scores, remove matched sentences. It repeats it until every sentence is paired.

High similarity detection uses 99.9% threshold for near-identical content:

- 'We reduced emissions by 15% in 2021' versus 'We reduced emissions by 15% in 2022' scores very high.
- 'Our renewable capacity increased significantly' versus 'We expanded solar portfolio substantially' scores much lower.

The final calculation works as follows:

- Document similarity scores range 0 to 1 for each method.
- High similarity ratio = (sentences above 0.999) ÷ total matched sentences.
- Average sentence similarity across all pairs.

Outputs for Communication Scoring

- Reporting Consistency dimension: High similarity ratios show recycled content while low ratios indicate fresh reporting effort. Sentence patterns reveal whether companies update their stories meaningfully or copy-paste from previous years.

This measures communication authenticity by separating genuine annual updates from lazy content recycling between reporting periods.

D.5.1 Similarity Verification

Only the SpaCy similarity results are shown in the verification below. TF-IDF and Jaccard scores are not included, since they produce document-level analysis without specific examples that are suited for clear visual. Although the 0.999 threshold may seem high, SpaCy's word vector similarity tends to assign high scores to related content especially within the same same company that discusses similar things in both years. As can be seen below the sentences around 0.95 often differ significantly in structure but still score high due to shared subjects. Keep in mind that the stop words are removed when comparing which increases this. The scores above 0.999 typically reflect near-identical text with only minor number changes, which validates choosing this threshold.

HIGH SIMILARITY (0.999 - 1.000) - Near identical/identical content

==

Found 5 pairs in this range. Showing top 3:

EXAMPLE 1 - SIMILARITY: 1.000000

2021: In relation to this regulation, it is calculated that, unrealized foreign exchange losses recognised under retained earnings/(losses) amounting to TL 2.236.188.968 and recognised under consolidated statement of profit or loss amounting to TL 2.970.58...

2022: In relation to this regulation, it is calculated that, unrealized foreign exchange losses recognised under retained earnings/(losses) amounting to TL 1.337.353.704 and recognised under consolidated statement of profit or loss amounting to TL 869.107....

——————————————————————————

EXAMPLE 2 - SIMILARITY: 1.000000

2021: With its product groups and 'advanced chemical coating material products', it is the solution partner of global brands, each of which is a leader in its field, in various sectors such as white goods, ceramic tiles, household and kitchenware, glass in...

2022: With its product groups and 'advanced chemical coating material products', it is the solution partner of global brands, each of which is a leader in its field, in various sectors such as white goods, ceramic tiles, household and kitchenware, glass i...

——————————————————————————

EXAMPLE 3 - SIMILARITY: 0.999774

*2021: * With the Communique of Ministry of Commerce issued on the official gazette dated 15 September 2018 regarding the regulation on loss of capital and excess of liabilities over assets in relation to Article 376 of Turkish Commercial Code numbered 6102...*

2022: With the Communique of Ministry of Commerce issued on the official gazette dated 15 September 2018 regarding the regulation on loss of capital and excess of liabilities over assets in relation to Article 376 of Turkish Commercial Code numbered 6102, ...

——————————————————————————

MEDIUM-HIGH SIMILARITY (0.95)

==

Found 11261 pairs in this range. Showing top 3:

EXAMPLE 1 - SIMILARITY: 0.949996

2021: We carried on incessantly with our corporate social responsibility activities with the accountability and commitment we feel for the environment and society

2022: Over time, it has realized hundreds of value-added projects with all its customers in the field of technology and has achieved many successes during this time.

————————————————————————————

EXAMPLE 2 - SIMILARITY: 0.949977

2021: Although 2021 was a dry year compared to previous years, we made good use of the profit margins in the Balancing Power Market with instant operations, especially thanks to the high water storage volumes of our Uluabat and Feke II Hydroelectric Power ...

2022: At Akenerji, with 1 wind and 7 hydroelectric power plants, which we gradually commissioned between 2009 and 2012, we have ensured that renewable energy sources are brought into the country's economy in a reliable, economical and high-quality manner.

————————————————————————————

EXAMPLE 3 - SIMILARITY: 0.949974

2021: As Turkey's leading electricity company, we continue to produce for our future with the understanding that we place the environment and people in our focus.

2022: As a company that focuses on power plant investments based on renewable resources in order to create resource diversity and cost advantage, we have raised the number of power plants with alternative energy sources over the years by carrying out many ...

————————————————————————————

E Communication Dimension Justification

The table below in this Appendix shows which output variables (coming from the NLP modules) belong to which communication dimension. The most important literature justification are presented in the table as well.

Table E.1: Communication dimensions and output variable justification

Dimension	Variable name	Justification (including references)
Green Communication Intensity	Green Term Frequency	Selective disclosure theory [140]; contextual frequency predicts information processing [35]
Green Communication Intensity	Vocabulary Diversity	Vocabulary diversity shapes stakeholder interpretation by indicating effort put into communication [180, 119]
Green Communication Intensity	Average Environmental Sentiment	Sentiment matters more than frequency for public impact [2]; receptivity to green communication affects green attitudes [248]
Green Communication Intensity	Renewable Energy Sentiment	Renewable energy carries distinctly positive sentiment that shapes public perception [135]
Green Communication Intensity	Climate and Emissions Sentiment	Climate topics carry distinctly positive sentiment [135]; environmental content coverage examining climate, energy, and sustainability [89]
Substantiation Weakness	Quantified Claim Intensity	Hard disclosure distinction [43]; quantitative disclosures associated with greater clarity and investor trust [156]
Substantiation Weakness	Evidence-Based Claim Intensity	Hard disclosure more credible [43]; robust, science-based evidence requirements [70]
Substantiation Weakness	Aspirational Claim Intensity	Soft disclosure distinction [43]; worse environmental performers exhibit more optimism [39]
Language Vagueness	Vague Language Intensity	Vague language creates scepticism [231, 84]; vague wording like "eco-friendly" means nothing concrete [209]
Language Vagueness	Hedge Language Intensity	Hedging expresses appropriate uncertainty for credible scientific communication [235]; distinguishing vague language from legitimate hedging [32]
Temporal Orientation	Future Orientation Ratio	Net-zero promises without recent progress [159]; future-focused communication patterns affect sustainability investment [137]
Temporal Orientation	Timeline Specificity	Hopeful messages need specific timelines instead of vague future promises [110]; timeline reference importance for credibility [11]
Reporting Consistency	Cross-Year Similarity Score	Systematic repetition of structure and content across years [41]; cross-year similarity analysis for detecting thematic patterns [120]
Reporting Consistency	High-Similarity Sentence Ratio	Sentence-level repetition analysis [120]; boilerplate language correlation with ESG rating problems [240]

F Communication Data and Scores

F.1 Required Subcomponent Data

This section presents the raw communication variables extracted from sustainability reports using the NLP modules described in Chapter 3.4.3. These subcomponents feed directly into the five communication dimensions used in the greenwashing risk assessment. These are the unnormalised values directly from the NLP modules before any scaling or transformation is applied to create the final communication dimension scores.

Table F.1: Communication Subcomponent Dataset

Company	Yr	GT Freq.	GT Uniq.	Avg Sent.	Ren. Sent.	Clim. Sent.	Qnt Int.	Evid Int.	Asp. Int.	Hed. Den.	Vag. Den.	Tl Pct	Fut. Rat.	TFIDF Sim.	Jac. Sim.	SpaCy Avg	SpaCy HS
Akenerji	21	2.46	1.89	0.18	0.48	0.01	17.98	16.01	18.79	0.85	3.23	7.50	0.22	0.96	0.71	0.97	0.36
Akenerji	22	2.16	1.97	0.14	0.41	0.00	16.32	14.81	22.53	0.88	3.16	6.38	0.34	0.96	0.71	0.97	0.36
Arendals	21	3.29	2.51	0.18	0.30	-0.11	10.98	11.09	23.13	0.90	4.75	6.25	0.38	0.93	0.59	0.95	0.22
Arendals	22	2.49	2.69	0.19	0.30	-0.10	11.63	12.84	22.72	1.16	4.35	4.05	0.39	0.93	0.59	0.95	0.22
Atlantica	21	3.60	2.48	0.03	0.25	-0.04	23.78	12.96	16.08	1.60	4.94	6.06	0.31	0.81	0.58	0.97	0.26
Atlantica	22	2.49	1.86	0.06	0.31	-0.16	29.42	9.21	18.05	1.83	3.79	4.41	0.25	0.81	0.58	0.97	0.26
CEZ	21	4.84	3.01	0.18	0.41	0.03	17.83	9.60	35.67	1.06	4.76	0.00	0.40	0.98	0.68	0.98	0.32
CEZ	22	4.20	2.84	0.14	0.38	0.03	14.04	12.83	28.34	0.87	4.72	5.13	0.32	0.98	0.68	0.98	0.32
EDF	21	6.72	4.42	0.34	0.36	0.15	15.07	5.37	22.09	0.62	4.96	0.00	0.44	0.44	0.28	0.89	0.00
EDF	22	8.60	3.82	0.52	0.44	0.13	16.38	12.50	33.62	0.72	6.00	0.00	0.61	0.44	0.28	0.89	0.00
EDP	21	4.99	2.31	0.26	0.47	0.05	16.21	11.09	24.89	0.93	5.61	3.56	0.43	0.84	0.39	0.95	0.02
EDP	22	2.41	1.45	0.08	0.36	0.04	17.38	12.71	23.10	1.17	3.76	4.18	0.31	0.84	0.39	0.95	0.02
ERG	21	5.31	2.58	0.27	0.43	0.07	22.37	10.74	25.71	0.97	5.55	5.66	0.41	0.97	0.61	0.96	0.10
ERG	22	5.29	2.49	0.26	0.38	0.04	21.34	13.92	22.27	0.92	5.43	2.11	0.38	0.97	0.61	0.96	0.10
Endesa	21	4.12	2.53	0.13	0.28	-0.02	17.67	10.86	27.31	1.10	5.32	3.65	0.36	0.99	0.68	0.98	0.27
Endesa	22	4.62	2.37	0.15	0.26	0.01	16.74	11.26	25.25	1.17	5.57	3.75	0.39	0.99	0.68	0.98	0.27
Ørsted	21	10.33	4.17	0.27	0.32	0.06	26.59	10.16	30.29	0.89	6.22	0.00	0.68	0.85	0.47	0.93	0.00
Ørsted	22	7.88	4.36	0.23	0.31	0.04	19.72	7.60	28.49	0.82	7.01	1.41	0.63	0.85	0.47	0.93	0.00
PGE	21	3.66	2.73	0.27	0.49	0.08	14.13	9.51	29.85	1.14	4.72	4.50	0.45	0.91	0.33	0.96	0.18
PGE	22	5.51	3.40	0.48	0.56	0.12	21.23	11.30	34.19	1.10	4.42	3.33	0.50	0.91	0.33	0.96	0.18
Romande	21	5.62	4.13	0.30	0.27	0.11	8.74	8.74	19.94	0.99	5.51	0.00	0.49	0.90	0.50	0.94	0.04
Romande	22	5.39	3.41	0.27	0.35	0.18	11.23	14.49	22.34	1.12	5.54	0.00	0.44	0.90	0.50	0.94	0.04
Scatec	21	5.46	4.23	-0.07	0.03	-0.02	13.31	13.10	30.04	0.56	6.09	0.00	0.79	0.92	0.61	0.96	0.11
Scatec	22	6.02	4.24	-0.04	0.13	-0.07	15.52	13.65	26.72	0.76	5.41	11.77	0.69	0.92	0.61	0.96	0.11
Solaria	21	5.31	3.47	0.23	0.40	0.08	11.63	11.85	19.41	1.11	4.83	0.00	0.30	0.89	0.50	0.96	0.01
Solaria	22	6.23	3.37	0.22	0.42	0.05	10.26	7.11	20.38	1.18	4.55	9.43	0.32	0.89	0.50	0.96	0.01
Terna	21	5.79	3.01	0.15	0.41	0.04	11.59	14.22	22.92	1.00	5.38	0.00	0.40	0.67	0.51	0.94	0.00
Terna	22	6.84	3.50	0.13	0.23	0.04	46.42	9.80	13.29	0.56	5.09	0.00	0.19	0.67	0.51	0.94	0.00

Variable definitions: GT_Freq. = Green Term Frequency (%), GT_Uniq. = Unique Green Terms Relative, Avg_Sent. = Average Sentiment Score, Ren._Sent. = Renewable Energy Sentiment, Clim._Sent. = Climate Emissions Sentiment, Qnt_Int. = Quantification Intensity Score, Evid_Int. = Evidence Intensity Score, Asp._Int. = Aspirational Intensity Score, Hed._Den. = Hedge Density, Vag._Den. = Vague Density, Tl_Pct = Commitment Timeline Percentage, Fut._Rat. = Future vs Past/Present Ratio, TFIDF_Sim. = TF-IDF Similarity, Jac._Sim. = Jaccard Similarity, SpaCy_Avg = SpaCy Average Similarity, SpaCy_HS = SpaCy High Similarity Ratio.

F.2 Dimension Scores

This section shows the normalised communication dimension scores (0-100 scale) calculated from the raw subcomponent data. These five dimensions represent the final communication measures that feed into the greenwashing risk ensemble analysis. Each dimension combines multiple subcomponents using the weighting structure detailed in Chapter 3.5.1, with scores normalised within each year to enable fair cross-company comparison.

Table F.2: Communication Dimension Scores

Company	Year	PCG	Subst. Weakness	Language Vagueness	Temporal Orientation	Reporting Consistency
Akenerji	2021	59.11	8.12	0.00	0.00	100.00
Akenerji	2022	44.45	30.69	13.01	19.11	100.00
Arendals	2021	21.01	71.83	45.41	23.51	72.75
Arendals	2022	25.40	53.53	31.51	39.92	72.75
Atlantica	2021	5.54	0.00	24.49	17.15	78.08
Atlantica	2022	55.25	42.98	0.00	12.98	78.08
CEZ	2021	43.32	100.00	39.38	58.95	97.03
CEZ	2022	59.43	66.48	47.97	21.89	97.03
EDF	2021	52.74	89.81	62.55	63.16	0.00
EDF	2022	59.80	81.06	80.55	100.00	0.00
EDP	2021	36.14	61.75	70.76	43.13	39.99
EDP	2022	0.00	47.39	18.11	24.93	39.99
ERG	2021	86.09	47.94	67.44	29.81	69.55
ERG	2022	11.23	29.72	62.17	48.00	69.55
Endesa	2021	8.54	66.44	55.59	35.27	93.57
Endesa	2022	42.24	66.00	58.35	41.48	93.57
Ørsted	2021	41.76	53.29	91.61	88.42	39.48
Ørsted	2022	56.38	100.00	100.00	96.43	39.48
PGE	2021	94.10	93.06	35.22	40.21	59.18
PGE	2022	100.00	85.19	34.83	63.32	59.18
Romande	2021	100.00	82.16	65.78	68.42	50.10
Romande	2022	69.71	39.99	59.08	69.50	50.10
Scatec	2021	0.00	75.06	100.00	100.00	66.49
Scatec	2022	32.51	53.17	66.46	54.55	66.49
Solaria	2021	45.41	53.56	39.75	48.42	52.41
Solaria	2022	42.55	93.91	35.31	0.00	52.41
Terna	2021	50.61	50.96	61.16	58.95	36.73
Terna	2022	83.61	0.00	65.22	24.64	36.73

Note: PCG = Performance-Communication Gap, Subst. = Substantiation.

These normalised scores reveal the final communication profiles used in the greenwashing risk assessment. Higher scores indicate greater potential concern within each dimension - for example, higher PCG scores suggest larger mismatches between environmental performance and communication intensity.

G Communication Correlation Analysis

This appendix examines how the communication dimensions in the GRAT correlate with each other. Using data from both years (n=28 observations), the analysis checks whether these dimensions measure different aspects of communication strategy or if they significantly overlap. These relationships are required for interpreting the multi-dimensional approach and the GRAT's theoretical structure.

Looking at the five communication-focussed dimensions (without performance influence), correlations show they are related but still capture different communication patterns. The average correlation is moderate (r = 0.497). Most pairs (8 out of 10) are statistically significant, which suggests companies coordinate their communication strategies instead of using completely separate approaches. Since no correlations go above 0.8, this supports using multiple dimensions rather than combining them into fewer measures.

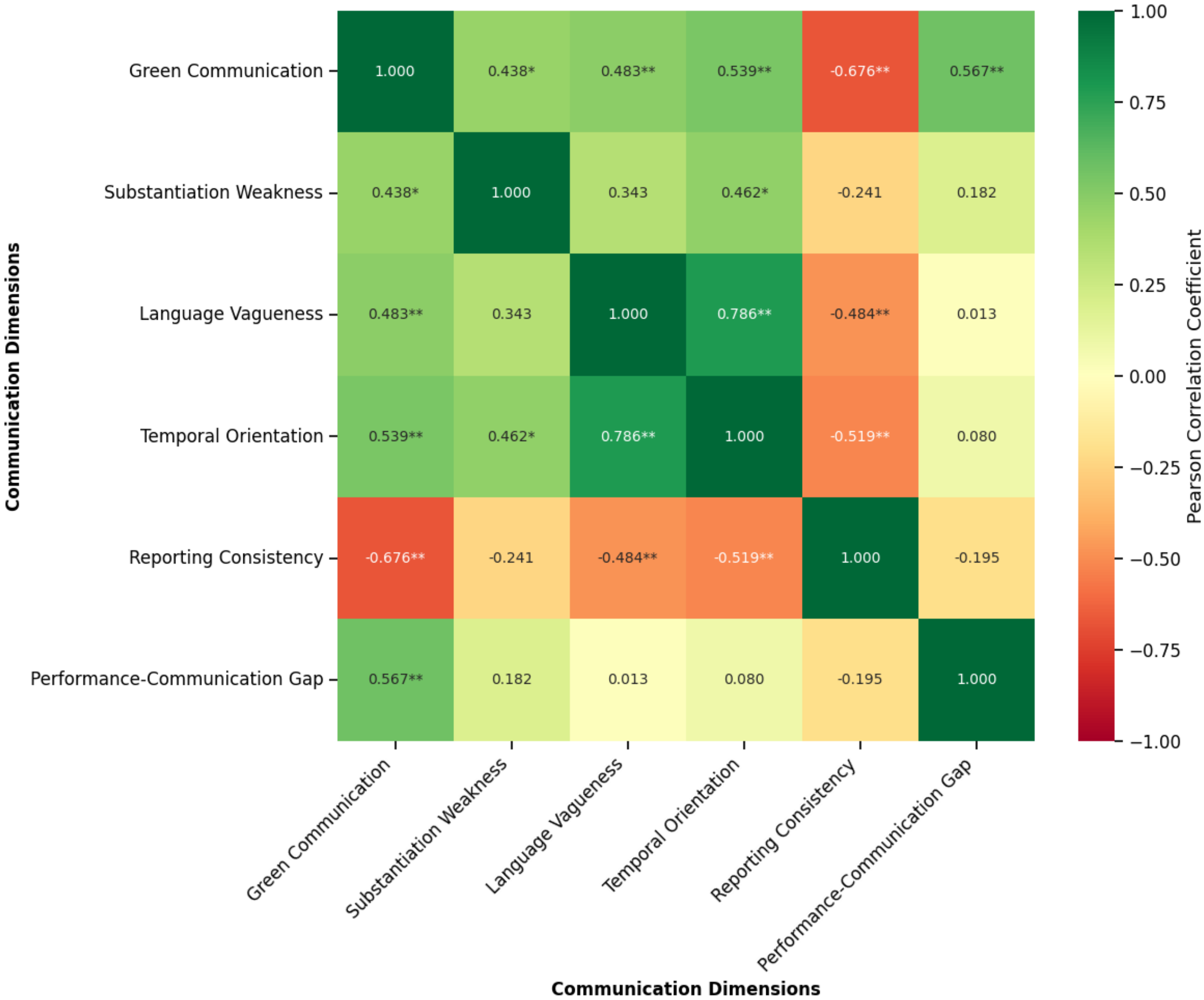

Figure G.1: Correlation Matrix of all Dimensions.

Note: ** $p \leq 0.01$, * $p \leq 0.05$

Figure G.1 shows the complete correlation matrix, including the PCG dimension. Language Vagueness and Temporal Orientation have the strongest connection ($r = 0.786$), companies that use vague language also tend to make future-focused claims without specific timelines. Reporting Consistency correlates negatively with most other dimensions, this could be because companies used different report structures during the pre-CSRD period when standardised formats were not required yet.

The PCG shows distinctly different correlation patterns compared to the communication quality dimensions. As expected, it correlates significantly with Green Communication ($r = 0.567$, $p < 0.01$) since this dimension directly contributes to calculating the absolute difference with performance scores. The PCG shows weak and non-significant correlations with the four communication quality dimensions ($r = 0.013$ to 0.182, all $p > 0.05$). While these non-significant correlations prevent strong conclusions, the weak correlations support the multi-dimensional GRAT design by suggesting that performance-communication mismatches and communication quality issues represent distinct greenwashing mechanisms that can occur independently. However, the limited sample size requires careful interpretation of whether these represent truly independent mechanisms.

Table G.1: Complete Communication Dimensions Correlation Matrix

Dimension Pair	Correlation (r)	p-value	Significance
Green Communication ↔ Substantiation Weakness	0.438	0.020	*
Green Communication ↔ Language Vagueness	0.483	0.009	**
Green Communication ↔ Temporal Orientation	0.539	0.003	**
Green Communication ↔ Reporting Consistency	-0.676	<0.001	**
Green Communication ↔ PCG	0.567	0.002	**
Substantiation Weakness ↔ Language Vagueness	0.343	0.074	ns
Substantiation Weakness ↔ Temporal Orientation	0.462	0.013	*
Substantiation Weakness ↔ Reporting Consistency	-0.241	0.216	ns
Substantiation Weakness ↔ PCG	0.182	0.357	ns
Language Vagueness ↔ Temporal Orientation	0.786	<0.001	**
Language Vagueness ↔ Reporting Consistency	-0.484	0.009	**
Language Vagueness ↔ PCG	0.013	0.946	ns
Temporal Orientation ↔ Reporting Consistency	-0.519	0.005	**
Temporal Orientation ↔ PCG	0.080	0.690	ns
Reporting Consistency ↔ PCG	-0.195	0.322	ns

*Note: , PCG = Performance-Communication Gap, ns = not significant, * = significant (p<0.05), ** = very significant (p<0.01)*

H Greenwashing Ensemble Statistics

This Appendix shows the ensemble statistics calculated across all 59,881 valid weight combinations for each company-year observation in the greenwashing risk assessment (see Table H.1). The ensemble method tests different weight allocations within theoretical constraints, creating a distribution of greenwashing risk scores for each case. The median scores become the final greenwashing risk values used throughout the analysis. Standard deviation and quartile ranges show how much scores change when weights vary.

Table H.1: Greenwashing Risk Ensemble Statistics

Company	Year	Mean	Median	Std Dev	Min	Max	Q25	Q75	IQR
Akenerji	2021	36.57	36.54	4.98	21.73	50.67	33.00	40.14	7.14
Akenerji	2022	40.78	40.55	3.83	30.87	53.59	37.84	43.55	5.71
Arendals	2021	43.67	43.66	2.86	35.52	52.60	41.64	45.68	4.04
Arendals	2022	40.95	40.73	2.25	35.22	48.60	39.30	42.49	3.19
Atlantica	2021	16.80	16.52	3.61	8.37	26.44	13.90	19.50	5.60
Atlantica	2022	42.93	42.98	3.62	31.02	53.31	40.43	45.49	5.06
CEZ	2021	65.85	65.80	3.19	56.67	76.49	63.56	68.07	4.51
CEZ	2022	59.83	59.78	3.02	51.73	70.00	57.70	61.90	4.20
EDF	2021	58.22	58.34	3.51	49.58	66.69	55.36	61.05	5.69
EDF	2022	65.38	65.61	4.24	51.88	75.93	62.27	68.61	6.34
EDP	2021	48.27	48.16	1.73	43.52	54.40	47.03	49.45	2.42
EDP	2022	22.32	22.56	2.25	15.79	27.97	20.72	24.00	3.28
ERG	2021	65.09	65.28	2.65	56.56	72.08	63.26	66.99	3.73
ERG	2022	33.91	33.82	3.21	25.00	43.21	31.50	36.25	4.75
Endesa	2021	42.86	42.71	4.08	32.27	54.26	39.88	45.74	5.86
Endesa	2022	56.51	56.26	2.51	50.38	65.02	54.63	58.23	3.60
Ørsted	2021	56.29	56.21	3.03	48.55	66.61	54.07	58.41	4.34
Ørsted	2022	75.74	75.71	3.21	66.34	85.25	73.44	78.05	4.61
PGE	2021	75.68	75.69	3.54	65.39	86.35	73.15	78.16	5.01
PGE	2022	78.68	78.82	3.29	69.46	87.68	76.31	81.12	4.81
Romande	2021	81.22	81.30	2.54	74.06	87.97	79.37	83.14	3.77
Romande	2022	58.37	58.33	1.38	54.57	62.46	57.36	59.34	1.98
Scatec	2021	51.95	51.99	5.69	38.40	67.34	47.59	56.24	8.65
Scatec	2022	48.85	48.83	1.92	44.24	54.20	47.37	50.29	2.92
Solaria	2021	48.03	48.03	0.60	46.49	49.75	47.58	48.48	0.90
Solaria	2022	50.77	50.62	3.48	43.79	62.12	47.96	53.29	5.33
Terna	2021	51.31	51.36	1.06	47.63	54.34	50.59	52.09	1.50
Terna	2022	46.96	46.59	3.80	38.16	57.96	44.06	49.65	5.59

Standard deviations range from high stability (0.60 for Solaria 2021) to moderate uncertainty (5.69 for Scatec 2021), reflecting the reliability of risk assessments across different weighting scenarios.

I (Sub)Component Sensitivity Analysis

This Appendix shows the results from sensitivity analysis that tested how robust the communication sub-component weights are in the GRAT. The analysis looked at 17 different scenarios where weights were adjusted by ±10% to see how this affected final greenwashing scores and company rankings (see *Table I1*). For substantiation quality components, smaller ±5% changes were used to keep the theoretical hierarchy intact (Evidence ≥ Aspirational > Quantified, as established by Clarkson et al. [43]).

Table I.1: Communication (Sub)Component Sensitivity Scenarios

(Sub)component	Baseline Weights	+10% Scenario	-10% Scenario
Combined Sentiment Score	0.6 (average sentiment), 0.2 (renew sent.), 0.2 (climate sent.)	0.66-0.17-0.17	0.54-0.23-0.23
Combined Green Term Score	0.7 (term frequency), 0.3 (vocabulary diversity)	0.77-0.23	0.63-0.37
Green Communication Integration	0.4 (comb. green term), 0.6 (comb. sentiment)	0.36-0.64	0.44-0.56
Substantiation Weakness*	0.30 (quantified), 0.35 (evidence), 0.35 (aspirational)	0.275-0.3625-0.3625	0.325-0.3375-0.3375
Language Vagueness	0.7 (vague), 0.3 (hedge)	0.77-0.23	0.63-0.37
Temporal Orientation	0.6 (future), 0.4 (timeline)	0.66-0.34	0.54-0.46
Reporting Consistency	0.7 (average similarity), 0.3 (high sim. sentences)	0.77-0.23	0.63-0.37
PCG Amplifier	1.5 (amplifier)	1.65	1.35

***±5% variation to preserve theoretical hierarchy**

In each scenario, all communication dimensions were recalculated with the modified weights. These results then went through the full ensemble methodology using all 59,881 valid weight combinations. This tests how sensitive sub-components are to changes and how they interact with system-level weight uncertainty.

Table I.2: Sensitivity Scenario Impact Rankings

Scenario	R_bar_S	avg_MAD	avg_CV	max_$\bar{R}_S$	impact_lvl
Amplifier -10%	0.214	0.625	0.834	1.40	Low
Reporting +10%	0.179	0.284	0.400	0.62	Low
Combined Sentiment -10%	0.143	0.210	0.530	0.62	Low
Green Term +10%	0.143	0.349	0.709	1.04	Low
Substantiation -5%	0.071	0.497	0.745	1.07	Low
Language Vagueness +10%	0.071	0.414	0.965	1.27	Low
Green Term -10%	0.071	0.351	0.704	1.10	Low
Green Communication -10%	0.071	0.715	1.657	1.73	Low
Temporal -10%	0.071	0.306	0.727	1.19	Low
Temporal +10%	0.071	0.315	0.747	1.21	Low
Amplifier +10%	0.071	0.565	0.754	1.26	Low
Language Vagueness -10%	0.071	0.440	1.030	1.41	Low
Substantiation +5%	0.000	0.496	0.747	1.10	Low
Green Communication +10%	0.000	0.698	1.618	1.72	Low
Combined Sentiment +10%	0.000	0.203	0.506	0.59	Low
Reporting -10%	0.000	0.282	0.399	0.61	Low

Note: R_bar_S = Average Ranking Stability; avg_MAD = Average Mean Absolute Deviation; avg_CV = Average Coefficient of Variation; max_$\bar{R}_S$ = Maximum Score Shift; impact_lvl = Impact Level

The $\bar{R}_S$ values show how stable rankings are when weights change. Values below 2.0 mean high robustness, while 2.0-4.0 suggests moderate stability, and anything above 4.0 indicates low robustness that needs attention [201]. All scenarios stayed well below 2.0, which means the GRAT is quite robust to these weight changes.

Table I.3: Individual Company Sensitivity Metrics

Company	Baseline Score	CV Pct	Score Range	MAD	avg_$\bar{R}_S$	Sens. Level
Atlantica	29.75	44.51	28.87	13.23	0.031	High
EDP	35.36	36.22	27.47	12.79	0.219	High
ERG	49.55	31.76	33.26	15.73	0.063	High
Romande	69.82	16.47	25.02	11.49	0.000	High
Ørsted	65.96	14.86	21.99	9.78	0.000	Moderate
Endesa	49.49	13.74	15.53	6.78	0.031	Moderate
EDF	61.98	5.95	9.45	3.65	0.000	Moderate
Akenerji	38.55	5.37	6.21	1.98	0.156	Moderate
Terna	48.98	5.10	7.70	2.39	0.063	Moderate
CEZ	62.79	4.87	7.99	3.01	0.063	Low
Arendals	42.20	3.69	4.96	1.47	0.156	Low
Scatec	50.41	3.36	5.43	1.59	0.063	Low
Solaria	49.33	2.89	4.35	1.31	0.250	Low
PGE	77.26	2.07	4.59	1.56	0.000	Low

Note: CV Pct = Coefficient of Variation Percentage; avg_ $\bar{R}_S$ = Average Rank Shift; Sens. Level = Sensitivity Level

Companies with CV values above 15% are highly sensitive to methodological choices. This reveals more complex communication patterns that change depending on how weights are adjusted. CV values between 5-15% show moderate sensitivity, while values below 5% point to straightforward communication strategies that stay consistent regardless of weight changes.

The high sensitivity companies (Atlantica, EDP, ERG, Romande) include both strong and weak performers. This shows that sensitivity comes from communication complexity, not performance quality. The most stable companies (PGE, Solaria, Scatec) keep consistent risk profiles across all scenarios. Their communication patterns are clearly defined and do not depend much on specific analytical choices. The known greenwashing companies from external validation (PGE, CEZ, Ørsted) all fall into the low to moderate sensitivity categories.

The impact levels combine $\bar{R}_S$ (average ranking shift) values with how much scores change. The amplifier factor caused the biggest ranking shifts ($\bar{R}_S = 0.21$), with reporting consistency changes coming second. Changes to substantiation quality and sentiment had much smaller effects on rankings.

This sensitivity analysis works with the ensemble methodology to increase transparency around methodological uncertainty while preserving theory-driven constraints. Complete ensemble and verification details appear in Appendices C, D, and H.

J Known Case Validation

J.1 Search Methodology

This appendix documents the systematic approach used to identify known greenwashing cases among the 14 sample companies. Documented accusations from credible sources provide validation benchmarks for GRAT testing.

Table J1 presents the multi-language search term framework, showing how company names combined with greenwashing terms across relevant languages.

Table J.1: Multi-Language Search Term Framework

Language	Base Terms	Greenwashing Terms	Combined Examples
English	Company name +	greenwash*, "misleading environmental claims", "false green claims"	"PGE" + greenwash*
Polish	Nazwa firmy +	"zielone pranie", greenwashing	"PGE" + "zielone pranie"
French	Nom d'entreprise +	écoblanchiment, verdissage	"EDF" + écoblanchiment
Spanish	Nombre empresa +	"lavado verde", "maquillaje verde"	"Endesa" + "lavado verde"
Danish/Norwegian	Firmanavn +	grønnvasking	"Ørsted" + grønnvasking
Czech	Název společnosti +	"zelené mytí"	"CEZ" + "zelené mytí"
Italian	Nome azienda +	"ambientalismo di facciata"	"ERG" + greenwashing
Portuguese	Nome empresa +	"lavagem verde"	"EDP" + "lavagem verde"
Greek	Όνομα εταιρείας +	πράσινο ξέπλυμα	"Terna Energy" + πράσινο ξέπλυμα
Turkish	Şirket adı +	"yeşil yıkama"	"Akenerji" + "yeşil yıkama"

Five primary source categories formed the foundation of this search strategy. NGO campaigns included Greenpeace national offices, ClientEarth, and Corporate Europe Observatory. These sources provided campaign pages, press releases, and legal actions. Platforms such as Danwatch, Demagog.org.pl, and CorpWatch provided environmental investigations and fact-checking reports. Industry publications including Energy Supply and sector trade press contributed business reporting on controversies. ESG assessment platforms like Sustainalytics and World Benchmarking Alliance provided controversy flags and risk assessments.

Table J.2: Primary Source Categories Searched

Source Type	Specific Platforms	Search Focus	Languages Used
NGO Campaigns	Greenpeace (all national offices), ClientEarth, Corporate Europe Observatory	Campaign pages, press releases, legal actions	Local + English
Investigative Journalism	Danwatch, Demagog.org.pl, CorpWatch	Environmental investigations, fact-checking	Local + English
Regulatory Bodies	consumer, financial, and competition regulators	Official complaints, enforcement actions	Local + English
Industry Publications	Energy Supply, sector trade press	Business reporting on controversies	Local + English
ESG Assessments	Sustainalytics, World Benchmarking Alliance	Controversy flags, risk assessments	English

January 1, 2020 to December 31, 2023 defined the temporal scope.

Inclusion Requirements:

- Specific accusations of misleading environmental communication
- Named accusers (NGOs and/or regulators)
- Claims related to sustainability reports or performance communication
- Verifiable source documentation

Exclusion Criteria Applied:

- Environmental incidents without communication allegations
- Marketing/sponsorship issues unrelated to performance-communication gaps
- Pre-2020 events (outside temporal scope)
- General industry criticism without company-specific claims

J.2 Search Results and Case Classification

Within the specified timeframe, four companies faced greenwashing accusations, although one case was later excluded based on scope criteria. Three companies received classification as known positive cases based on documented accusations. All remaining companies showed clean records across searched sources.

Table J.3: Search Results Summary

Company	Accusations Found	Source Types	Classification
PGE SA	4 documented cases	NGO campaigns, fact-checking, legal proceedings	Known Positive Case
CEZ Group	2 documented cases	NGO reports, fact-checking analysis	Known Positive Case
Ørsted	1 documented case	Regulatory complaint, investigative journalism	Known Positive Case
Endesa SA	2 documented cases	NGO reports, journalism investigations	Clean Record (see below)
EDF	Limited evidence	Historical cases outside timeframe	Clean Record but historical reference
All others (9 companies)	No cases found	Clean search results across all sources	Clean Record

Among all companies examined, PGE (Polska Grupa Energetyczna SA) attracted the most greenwashing accusations during the study period. Greenpeace Poland filed a climate lawsuit in March 2020 demanding a zero emissions strategy by 2030. Court hearings continued through 2022 [95]. Between 2021 and 2023, campaigns accused PGE of presenting itself as a "green transformation leader" despite generating just 4% renewable energy [97, 96]. A 2023 investigation by Demagog [50] found that PGE's emissions rose 19% between 2020 and 2021, contradicting its green messaging. Over six years, the company allocated only 3.5% of capital expenditure to renewable development [50].

Accusations against CEZ Group came from Greenpeace Czech Republic [94] in September 2023. The organization challenged CEZ's claims of environmental friendliness while aiming to continue Bílina coal mining operations until 2035 [94]. This mine produces 8 million tonnes of coal annually, representing 10% of Czech Republic's total emissions. An earlier review by independent experts concluded that 20 of the 25 claims in CEZ's "Nuclear Seven" letter to the EU Commission were false or misleading about nuclear power's environmental impact [58].

In March 2023, a formal complaint was filed against Ørsted. Danwatch [47] and the Council for Green Transition filed a complaint with the Danish Consumer Ombudsman over Ørsted's marketing of biomass burning as "sustainable" [47, 59]. Professor Jannick Schmidt concluded that Ørsted's calculation methods "hide the most serious climate impacts" of biomass use. About 75% of Denmark's renewable energy mix is made up of biomass, with Ørsted as the largest user [47]. According to the Danish Consumer Council, the practice violates legal standards for marketing.

Cases involving Endesa SA were found but excluded due to scope limitations. CorpWatch [45] documented "academic greenwashing" through university partnerships [45]. Multiple Spanish outlets reported COP25 sponsorship controversies [44, 146]. These cases were excluded as they dealt with sponsorship and academic ties, rather than emissions or misstatements in sustainability disclosures. COP25 events (in 2019) also fell outside the 2020-2023 timeframe. EDF showed none to limited evidence within the timeframe, but historical cases exist from 2015-2016 [215].

No documented greenwashing accusations appeared for the remaining nine companies during the search across all source categories (as seen in *J.3*). Several companies appeared in positive sustainability contexts instead. World Benchmarking Alliance [244] assessments featured EDP (Energias de Portugal) with 74% renewable generation by 2022 [244]. Strong ESG ratings went to Atlantica Sustainable Infrastructure PLC, with Sustainalytics [224] rating their risk management as "Strong" with no controversy flags [224]. Norwegian Consumer Authority guidelines requiring adequate documentation for green claims govern Scatec ASA operations [114].

Multiple independent sources provided cross-checking where possible. Validation groups for discriminant testing were formed based on identified cases: known positives (n = 3; PGE, CEZ, Ørsted) and companies with clean records (n = 11). By introducing this classification, MWU testing can be used to test whether the GRAT separates accused companies from those without known greenwashing incidents.

J.3 Mann-Whitney U Testing Results

Whether companies with known greenwashing accusations scored higher on the GRAT than companies with clean records was tested using the MWU test. No assumptions about normal distributions are required for this test, making it appropriate for small sample validation studies [243, 143, 171]. However, tests like these with only 14 observations present serious statistical limitations due to central limit theorem requirements. These results are indicative only and should be interpreted with extreme caution. The observed patterns could represent genuine GRAT accuracy, or they could be random phenomena that occurred between 2021-2022. More observations would give far better results, but these represent all available companies with complete data given our constraints.

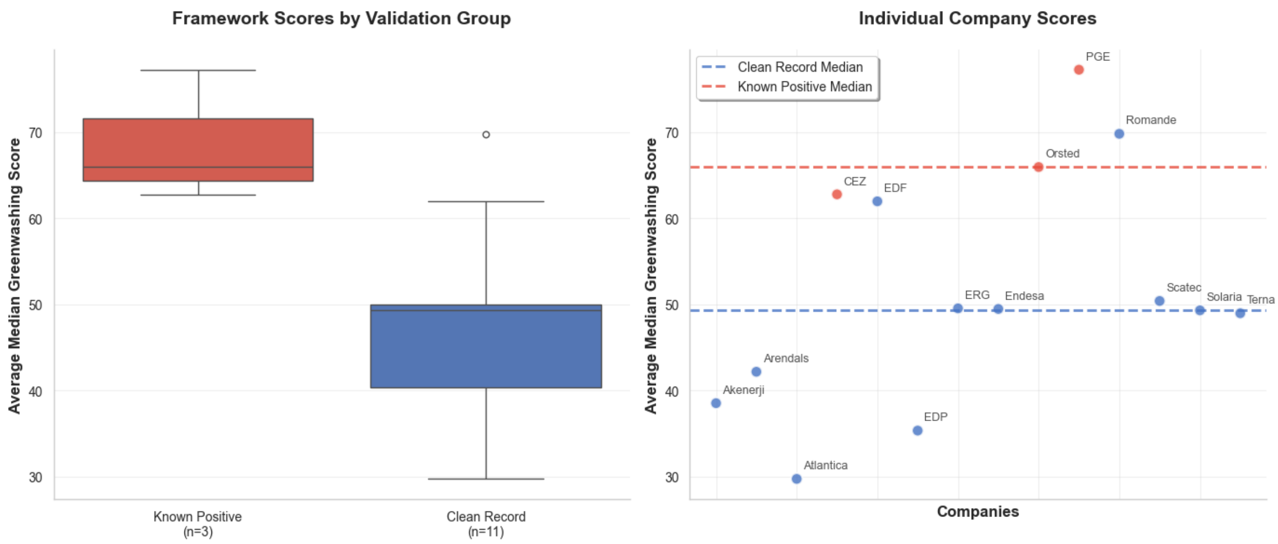

Figure J.1: Known Case Validation Results

As can be seen in *J.1* above, all three known positive cases ranked in the top four positions out of 14 companies (positions 1, 3, and 4), and every known positive case scored above the clean record median. The results suggest that the GRAT may potentially capture group differences and identify companies involved in confirmed greenwashing cases, although this pattern could also reflect random variation given the statistical limitations.

Table J.4: Known Case Validation Statistics

Test Parameter	Value	Interpretation
Known Positive Cases	n = 3	PGE (77.26), CEZ (62.79), Ørsted (65.96)
Clean Record Companies	n = 11	All remaining companies
U-statistic	31.0	Test statistic
p-value (one-tailed)	0.0110	Statistically significant (p < 0.05)
Effect Size (r)	-0.879	Large effect
Known Positive Median	65.96	SD = 7.61
Clean Record Median	49.32	SD = 11.40
Difference in Medians	16.64	Known positive cases score higher
Known Positive Mean	68.67	SD = 7.61
Clean Record Mean	47.76	SD = 11.40
Difference in Means	20.91	Known positive cases score higher

All scores were ordered from lowest to highest, and rank sums were computed per group. Positive cases (R_1): PGE (77.26), CEZ (62.79), and Ørsted (65.96) tended to rank above companies with no known accusations. Standard Mann-Whitney U-statistic calculations followed:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2} \tag{J.1}$$

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2} \tag{J.2}$$

$$r = 1 - \frac{2U}{n_1 \times n_2} \tag{J.3}$$

where:

- U_1 = U-statistic for known positive cases (group 1)
- U_2 = U-statistic for clean record companies (group 2)
- R_1 = sum of ranks for known positive cases
- R_2 = sum of ranks for clean record companies
- n_1 = number of known positive cases (n = 3)
- n_2 = number of clean record companies (n = 11)
- U = smaller of U_1 and U_2 used for effect size calculation
- r = rank-biserial correlation coefficient measuring effect size

With U = 31.0, $n_1 = 3$, and $n_2 = 11$, this results in:

$r = 1 - \frac{2\times 31.0}{3\times 11} = 1 - \frac{62}{33} = -0.879$

Clean record companies tend to rank lower than known positive cases, indicated by the negative value. With $|r| \geq 0.5$ considered a large effect by standard criteria, the observed value of 0.879 clearly represents a large effect. Statistical testing rejected the null hypothesis of no difference between groups (p = 0.0110 < 0.05), but again, the MWU test has very low statistical power with such a small sample size. Known positive cases scored substantially higher than clean record companies, with a mean difference of 20.91 points. This finding suggests the GRAT may potentially identify companies with documented greenwashing patterns, although definitive conclusions cannot be drawn given the sample limitations.

As a side observation, two out of three companies mentioned in a positive sustainability context ended up with the two lowest greenwashing scores.

K Reproduction Resources

The complete GRAT implementation is available through a GitHub repository containing all code modules and documentation necessary for GRAT replication and adaptation to new datasets.

Link: https://github.com/AadB010/Greenwashing-Risk-Assessment-Tool

Software Requirements
All technical dependencies are specified in the requirements.txt file included in the repository. The implementation uses Python with standard data science libraries including pandas, numpy, spaCy, transformers, and scipy for statistical analysis.

Repository Structure
The analysis follows a sequential notebook structure organised into five main categories:

01_data_extraction/: Example modules demonstrating CDP climate data processing, Refinitiv Eikon integration, and data harmonisation. These files illustrate the data preparation approach but are not required to follow this exact structure. Users can adapt these methods to their specific data sources and formats.

02_performance_analysis/: Environmental performance scoring implementation using constrained ensemble methodology across four core components with sector-specific adjustments.

03_communication_analysis/: Text processing and multi-dimensional NLP analysis including PDF to text extraction and cleaning using SpacyLayout, followed by five analysis modules for green communication intensity, substantiation weakness, language vagueness, temporal orientation, and reporting consistency.

04_integration_scoring/: Performance-Communication Gap calculation and final risk scoring using ensemble methodology across 59,881 valid weight combinations.

05_validation/: External validation against documented cases and sensitivity analysis across 17 weight variation scenarios.

Data Limitations and Requirements
Important Note: The code will not run as-is because it requires specific data files that are not included in the repository. Users must obtain their own data sources and adjust the code to extract relevant metrics from their datasets.

Performance Analysis Requirements:

- Scope 1 and 2 emissions data (self-reported and third-party verified)
- Emission intensity metrics with consistent denominators
- Target documentation (reduction percentages and target years)
- Renewable energy intensity using consistent denominators
- Minimum two consecutive years of data

Communication Analysis Requirements:

- English-language sustainability reports in digital format
- Aligned reporting periods across analysis years
- PDF documents suitable for text extraction using SpacyLayout

Execution Guide
Detailed implementation guidance is provided through multiple sources. The repository README file contains overview and setup instructions. Markdown cells within each notebook module provide step-by-step explanations of calculations and methodology. Chapter 5 of this thesis offers practical application guidance including data preparation, execution sequence, and result interpretation.

Execute notebooks sequentially by folder numbering, ensuring data requirements are met before proceeding to subsequent analysis stages. Each module includes validation steps and produces ensemble statistics suitable for uncertainty analysis and methodological robustness testing.

GRAT's modular design enables adaptation to different sectors by modifying performance metrics, communication terminology, and weight hierarchies while maintaining the core methodological approach.