

Cross-Industry ESG Analysis

Uncovering Sustainability Sentiments, Greenwashing & Corporate Responsibility Insights Through Natural Language Processing

~Team 8, MSBA Group A

Why ESG Matters in Business?

The way businesses respond to today's sustainability challenges is under the spotlight like never before. Companies are not just measured by profits, but they are judged on how they *treat the world around them and everyone they impact*.

Environmental, Social and Governance (ESG) frameworks, now, set the benchmark for responsible business. The **environmental** pillar is all about a **company's effect on nature** - carbon emissions, resource use, and climate action. **Social** focuses on how well companies **care for people**, from employees and customers to wider communities. **Governance** covers how **businesses are led and managed**, with attention to transparency, ethics, fair practices and accountability. Whether it's **attracting investment, earning consumer trust or staying ahead of future regulations**, excelling across all three ESG pillars is now essential for success.

Analysis Objectives:

- Identification of the industry-specific sustainability focus:** Which sustainability issues stand out in each industry and how do companies emphasize different ESG pillars in their reporting?
- Spotting potential greenwashing:** Are some companies talking a bigger sustainability game than they're actually playing? By digging into the language of their reports, this study seeks to spot the difference between bold claims and credible action.
- Comparing company and external views:** What do outsiders (journalists, analysts and the public) really think compared to what companies say? By comparing news sentiment to official ESG disclosures, the project uncovers gaps and hints of "greenwashing" across sectors.

Workflow Pipeline: Data Collection & Methodology

Data Source

Studied leading companies across 5 industries: **Oil & Gas, Agriculture, Fashion/Retail, Pharmaceuticals and Healthcare**.

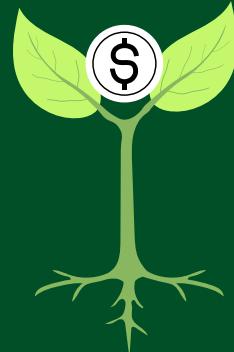
Company **sustainability reports** were collected directly from **publicly available PDFs**. **News articles** were gathered from the **web** to capture what's trending in media sentiment.

ESG Pillar Classification

To highlight the ESG pillar each company prioritized, **TF-IDF** was used to identify the **jargon most unique to each report**. Custom lists of pillar-specific keywords, **expanded** and **fine-tuned** using **WordNet**, made sure the results genuinely **reflected industry realities**, catching not just "emissions" but also terms like "methane" for Oil & Gas or "clinical trials" for Pharmaceuticals.

Detecting Greenwashing

Rather than just counting positive or negative words, the study **compared** how many times **aspirational phrases** like "committed to" or "vision" appeared versus **concrete action words** like "implemented" or "reduced". If companies **leaned** too much on **lofty goals**, it raised a **red flag for potential greenwashing**.



Text Extraction & Cleaning

Text extraction involved converting **corporate sustainability reports** from PDFs into **clean, structured text** suitable for analysis. Company ESG-related **news headlines and articles** were gathered through **web scraping** and **APIs**, capturing timely market sentiment. **Preprocessing** steps included **removing irrelevant corporate jargon, stopwords and formatting noise**, normalizing and standardizing word usage.

Sentiment Analysis

For both corporate reports and news headlines, specialized tools assessed the **"emotional tone"** and how **subjective or objective** the language was. **VADER** was used for the concise, punchy language found in headlines, while **TextBlob** handled more in-depth reports, **scoring** both **positivity/negativity and subjectivity**.

Word Clouds

Word clouds were created after careful **cleaning, standardizing, lemmatizing and ranking of words** using **TF-IDF** to ensure balance and clarity. Only the **most meaningful and representative words** for each ESG pillar were given **prominence**, so each cloud truly reflects the **sector's ESG priorities and overarching industry goals**.

Result 1: Each Industry's Dominant ESG Pillar & VADER Scores

Industry	Environmental	Social	Governance	Dominant Pillar
Agriculture	5	4	1	E
Fashion/Retail	3	12	0	S
Healthcare	0	10	0	S
Oil & Gas	7	2	0	E
Pharmaceuticals	1	7	1	S

Insights:

- Agriculture and Oil & Gas: Environmental pillar.**
- Healthcare, Pharmaceuticals & Fashion/Retail: Social pillar.**
- ESG reporting in *Fashion/Retail* centers on **labor rights** and in *Healthcare*, on **patient care and community wellbeing**.
- VADER sentiment analysis** revealed **consistently high positivity** in sustainability reports across all industries, showing a **strong tendency toward aspirational language**.
- This **uniform "positive image"** suggests that companies often **accentuate promotional messaging**, which can **overstate actual ESG performance**, even in sectors with controversy or risk.

Result 2: Greenwashing Risk

Insights:

- Aspirational Language vs. Concrete Action Words.
- Pharmaceuticals and Oil & Gas show the **highest risk**, with 78% of companies in each labeled high-risk and **average scores of 8.07 and 6.38**, indicating ongoing structural challenges.
- Healthcare is close behind, with 70% of companies flagged as **high risk**; Fashion/Retail and Agriculture display more **moderate risk profiles**.
- Every industry includes both low-risk and high-risk companies**, demonstrating that greenwashing is a widespread concern, not limited to any one industry.

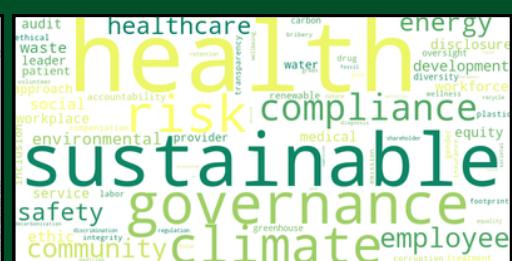
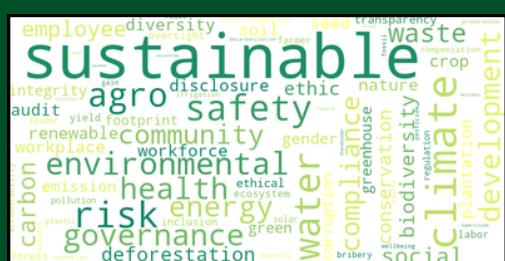
Industry	Average of Risk Score	Percentage of Companies at High Risk
Agriculture	6.15	50%
Fashion/Retail	5.94	53%
Healthcare	6.35	70%
Oil & Gas	6.38	78%
Pharmaceuticals	8.07	78%



Result 3.1: ESG Pillar Word Clouds

Insights:

- Environmental Pillar:** Dominated by themes like **sustainability, water, energy, climate, carbon and emissions**, reflecting a strong industry focus on **climate action and resource management**.
- Social Pillar:** Priorities are centered on **health, safety and community**, along with **inclusivity and employee wellbeing**, highlighting a **shared commitment to safe and healthy workplaces and social value**.
- Governance Pillar:** Features words like **governance, compliance, risk, audit and transparency**, showing a clear emphasis on **regulatory alignment and ethical practices across companies**.



Agriculture Word Cloud

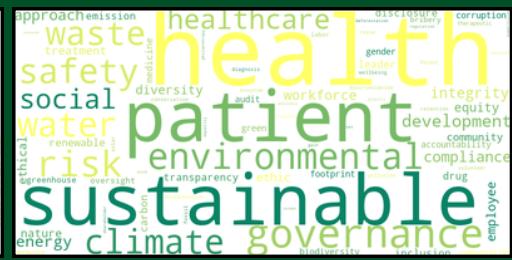
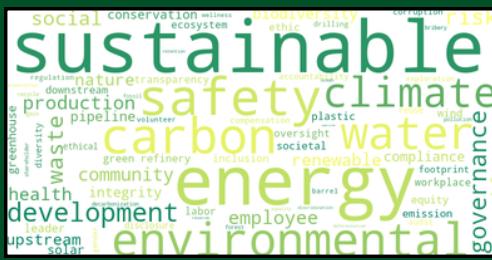
Fashion/Retail Word Cloud

Healthcare Word Cloud

Result 3.2: Industry Word Clouds

Insights:

- Agriculture:** Emphasizes on **Environmental** and **Social** aspects ("sustainable", "climate", "community" & "workforce").
- Fashion/Retail:** Dominated by **Social** aspect ("compliance", "wellbeing" & "workforce") with focus on **labor ethics and workplace culture**.
- Healthcare:** Centers on **Social** welfare themes ("sustainable", "health", "risk" & "community").
- Oil & Gas:** Focuses on **Environmental** transition ("sustainable", "environmental", "climate" & "energy") with **attention to workforce and community health**.
- Pharmaceuticals:** Highlights **Social** priorities ("patient", "health" & "development") with focus on **patient safety and health equity**.



Oil & Gas Word Cloud

Pharmaceuticals Word Cloud

Industry	Weighted Mean Delta Polarity	Weighted Mean Delta Subjectivity
Agriculture	0.018	-0.044
Fashion/Retail	0.004	-0.048
Healthcare	-0.002	-0.026
Oil & Gas	-0.015	-0.066
Pharmaceuticals	-0.015	-0.026

Result 4: News Sentiment Analysis

Insights:

- Agriculture:** Highest positive delta polarity, suggesting greater openness to external critique.
- Oil & Gas:** Lowest, indicating more transparency in reporting.

Company Specific Outliers

- Kuala Lumpur Kepong (Agriculture) and Pfizer (Pharmaceuticals) emphasize **governance** over their industry's typical ESG pillar.
- Marks & Spencer, Levis and Ralph Lauren (Fashion/Retail) stand out for **environmental focus** and **low controversy**, signaling mature ethical practices.
- In Oil & Gas, Exxon Mobil and Marathon Petroleum highlight **social themes** (e.g. workforce, employee welfare) over environmental, reflecting industry-specific priorities.

Key Takeaways & Implications

- Companies often **favor aspirational over concrete language** in reports, inflating their ESG image versus real action.
- NLP-driven analysis highlights **industry differences in pillar priorities** and **greenwashing risk**, with **word and sentiment analysis** revealing both **strengths and gaps**.
- Insights help **investors, regulators, analysts and executives** better evaluate **authenticity in ESG reporting** and guide **more informed decisions**.

News Media VS Company Reporting

- Agriculture** shows the **largest positive polarity gap**, with companies presenting themselves **more positively than media**, suggesting **higher greenwashing risk**.
- Pharmaceuticals** also show a **slight positive gap**, while **Oil & Gas** and **Fashion/Retail** have **negative gaps**, indicating **more cautious or restrained reporting**.
- Healthcare's polarity gap** is nearly **neutral**, reflecting **balanced messaging** between corporate and media sources.