

FROM ESG NARRATIVES TO EXPLAINABLE RISK SIGNALS: AN AI-DRIVEN FRAME ANALYSIS FRAMEWORK FOR SUSTAINABLE FINANCE

Nadiia Novytska¹, Ilias Sviatko¹, Jorge Augusto Meira¹, Timur Erkimbaiev¹, Radu State¹,
Alex Panican², Esli Spahiu²

¹ SnT, University of Luxembourg, Luxembourg

² Luxembourg House of Financial Technology, Luxembourg

E-mail: nadiia.novytska@uni.lu

Abstract: Corporate ESG disclosures are increasingly used in financial decision-making, yet their analytical value is constrained by narrative complexity and the risk that forward-looking commitments are read as evidence of delivered performance. This study demonstrates an explainable AI-driven framework that operationalises framing theory at claim level and links narrative classification to traceable report-level indicators. A theory-grounded 12-frame taxonomy distinguishes claim posture across epistemic, agency, valence and legitimacy families. Six corporate reports are processed through a structured large-language-model workflow and presented in an interactive dashboard linking indicators to claim-level evidence. The cases reveal distinct commitment-delivery and disclosure-credibility profiles, including comparatively stronger profiles for BNP Paribas, Deutsche Bank and Société Générale and more promise-heavy profiles for Shell, DWS and Nestlé. The signals support sustainable-finance screening and evidence triage but do not verify ESG performance or establish greenwashing.

Keywords: ESG reporting, framing analysis, greenwashing screening, natural language processing, sustainable finance

INTRODUCTION

Environmental, social and governance (ESG) disclosures increasingly influence investment, lending, regulatory and assurance decisions. Yet greater disclosure volume does not necessarily improve clarity or comparability. Sustainability reports remain long, heterogeneous, and narrative, and forward-looking commitments may be interpreted as evidence of performance delivered. Analytically important distinctions are often encoded in wording: whether a company reports a realised result, announces a future target, acknowledges risk, attributes responsibility or invokes external authority. Conventional ESG assessment captures reported performance but provides limited insight into how claims are constructed and legitimised. This limitation matters because ESG ratings diverge substantially across providers (Berg et al., 2022), while greenwashing may operate through selective disclosure, vague claims and impression management rather than only through demonstrably false statements (Delmas & Burbano, 2011; Marquis et al., 2016).

Framing theory provides a basis for analysing these communication choices. Goffman (1974) conceptualised frames as interpretive structures through which events are understood, while Entman (1993) defined framing as the selection and increased salience of particular aspects of reality. Applied to ESG reporting, the same topic may be framed as delivered progress, a future commitment, a material risk, a commercial opportunity or a claim supported by recognised authority. Existing computational studies commonly focus on topics, sentiment, keyword frequency or entities, offering less capacity to distinguish claim posture, agency, materiality and evidential support or to connect claim-level classifications with traceable report-level indicators. Domain-specific financial language further reinforces the need for grounded taxonomies (Loughran & McDonald, 2011).

This study develops and demonstrates an explainable AI-driven framework that operationalises framing theory at claim level and connects ESG narrative classification with traceable report-level indicators and role-specific sustainable-finance views. It asks how a fixed taxonomy can represent the principal communication functions of ESG claims, how claim classifications can be converted into transparent report-level indicators and professional views, and what value and limitations emerge from an operational prototype. The study adopts an applied multidisciplinary research design combining sustainable-finance analysis and framing theory with large-language-model-based information extraction and interactive visualisation. A design-science approach is used to develop and demonstrate the analytical prototype. The results reveal differences in promise-delivery balance, disclosure credibility and external-authority signals,

while confirming that narrative analytics should support human review rather than operate as an autonomous greenwashing verdict.

METHODOLOGY

The study follows a design-science orientation, in which an information-system artefact is constructed to address a defined analytical problem and assessed through demonstration, transparency and a planned validation programme (Hevner et al., 2004; Peffers et al., 2007). The 12-frame taxonomy was developed deductively through theory-to-operation mapping rather than induced from the six-case corpus. Candidate frames were retained when they represented a distinct communication function in framing, legitimacy, stakeholder or sustainability-reporting research; could be identified from observable claim-level evidence; and supported a relevant sustainable-finance distinction. The taxonomy comprises four families. The epistemic family separates Progress, Commitment and Aspiration. The agency family distinguishes Proactive agency, Deflection and Stakeholder salience. The valence family covers Risk/threat, Opportunity and Materiality signal. The legitimacy family differentiates Compliance, External authority and Narrative continuity. The conceptual model permits one dominant frame and relevant secondary frames, while the current dashboard displays one dominant frame per semantic triple. The taxonomy remains subject to expert validation.

The demonstration corpus follows a contrastive purposive case-selection strategy combining institutional relevance and controversy salience. BNP Paribas, Deutsche Bank and Société Générale were selected for their relevance to the Luxembourg financial-technology ecosystem and potential for domain-informed feedback. Shell, DWS and Nestlé were selected as controversy-salient stress-test cases. Awareness of the controversies informed case selection, but allegations, enforcement findings and external assessments were not supplied to the model or used in indicator calculations. All six reports were uploaded and processed through the same workflow.

The unit of analysis is a claim-bearing sentence or a claim extracted from a sentence. The prototype uses a large language model as a constrained information-extraction and classification component: it assigns claims within the predefined 12-frame taxonomy and structured-output schema rather than generating categories. Retained claims are converted into Subject-Predicate-Object semantic triples and enriched with a dominant frame, realised or hedged status, confidence score, rhetorical appeal, Entman function and, where relevant, an invoked reporting standard and standards-deviation category. The conceptual classification supports multi-frame interpretation when a sentence performs several communication functions; the present interface operationalises the dominant frame and retains secondary-frame implementation for subsequent validation. The rhetorical classification draws on ethos, logos and pathos in sustainability reporting (Higgins & Walker, 2012). This follows the information-extraction logic of converting unstructured text into explicit records that can be stored, queried and aggregated (Conlon et al., 2008), while interactive visualisation preserves links between report-level signals and source evidence (Mitri, 2022).

Claim-level records are aggregated into explainable risk signals and frame distributions. The Commitment-Achievement ratio compares forward-oriented communication with realised progress; the inverse Walk-versus-Talk ratio expresses delivery relative to promises. Additional indicators include Positivity Skew, the share of claims invoking F11 External authority and an illustrative Disclosure Credibility Index combining selected positive and negative frame signals. The index weights and thresholds are design parameters rather than statistically estimated coefficients. The dashboard organises the same evidence for sustainable-fund review, credit analysis, portfolio management and risk-and-compliance assessment. The architecture also supports sustainability-linked bond and loan monitoring by comparing targets with subsequently reported outcomes. Year-on-year Frame-Drift Trajectory and sector comparison remain planned modules and are not treated as empirical outputs in the present demonstration.

The prototype is evaluated as a design demonstration rather than a validated classifier. A subsequent validation protocol should use expert-annotated corpora, inter-rater agreement, precision, recall, F1 scores, sensitivity analysis, multilingual and adversarial robustness tests, and user studies comparing dashboard-supported review with conventional manual analysis.

FINDINGS

The prototype successfully processed the six reports and rendered linked analytical views. The indicators operate as explainable risk signals rather than reputational scores. BNP Paribas, Deutsche Bank and Société Générale show comparatively more balanced narrative-disclosure profiles (DCI 66-74; C/A 0.97-1.32; W/T 0.76-1.03), while Shell, Nestlé and DWS remain more promise-heavy (DCI 40-57; C/A 1.73-2.05; W/T below 0.60). Positivity Skew adds a distinct dimension: Deutsche Bank and Shell show balanced positive-to-negative evidence mixes, whereas BNP Paribas, Société Générale, Nestlé and DWS show high positivity skew. BNP Paribas records the strongest overall credibility and commitment-delivery profile (DCI 74; C/A 0.97; W/T 1.03), despite a low F11 External authority share (1%) and high Positivity Skew (4:1). Deutsche Bank has the highest F11 share (2%). Nestlé shows the largest commitment-delivery imbalance, while DWS records the lowest DCI. Table 1 summarises the outputs.

Table 1. Comparative dashboard outputs across six corporate reports

Company	Year	DCI	C/A	W/T	F11	Positivity skew
BNP Paribas	2023	74 - Solid	0.97 - Low	1.03	1%	4↑ : 1↓ - high
Deutsche Bank	2025	67 - Adequate/Watch	1.08 - Moderate	0.93	2%	5↑ : 2↓ - balanced
Société Générale	2024	66 - Adequate/Watch	1.32 - Moderate	0.76	1%	3↑ : 0↓ - high
Shell plc	2025	57 - Weak	1.73 - Elevated	0.58	1%	4↑ : 2↓ - balanced
Nestlé	2025	46 - Weak	2.05 - High	0.49	1%	3↑ : 1↓ - high
DWS	2025	40 - Weak	1.75 - Elevated	0.57	1%	3↑ : 1↓ - high

Reference: DCI: Disclosure Credibility Index; C/A: Commitment-Achievement ratio; W/T: Walk-versus-Talk ratio; F11: External authority share; Positivity Skew: reports the positive-to-negative evidence count and its screening category.

The results demonstrate why aggregate indicators must remain linked to claim-level evidence. A quantified reduction with a baseline is classified as Progress, whereas a future reduction target is classified as Commitment. A realised sustainable-finance total may still receive a standards-deviation flag if its scope or product taxonomy is insufficiently specified. Similarly, acknowledgement of physical climate risk may be classified correctly as Risk/threat but flagged as under-disclosed when the exposure is not quantified. External assurance is recorded as External authority, but it does not establish the factual accuracy of every reported KPI.

Sustainable-fund and credit users can use commitment-delivery ratios, materiality links, external-authority signals and standards-deviation categories to prioritise claims for supporting evidence or covenant review. The records also support sustainability-linked finance monitoring by distinguishing announced targets from subsequent achievements. Portfolio and compliance users can examine omissions, legitimacy strategies and source links to identify changes in narrative posture or areas where compliance language substitutes for outcome evidence. These functions support evidence triage and escalation but do not determine fund inclusion, contractual breach or legal compliance.

Post-hoc external contextualisation showed directional consistency between the prototype's weaker profiles and documented regulatory or consumer controversies concerning DWS, Shell and Nestlé (Advertising Standards Authority, 2023; European Consumer Organisation, 2023; U.S. Securities and Exchange Commission, 2023). Because controversy awareness influenced case selection, this comparison is illustrative rather than validating. The sample is small, purposive and heterogeneous across sectors and years. The classifications have not yet been tested against an expert-annotated gold standard, and the indicator weights and thresholds remain uncalibrated.

CONCLUSIONS

The study shows that ESG frame analytics can convert corporate narratives into explainable risk signals without equating linguistic credibility with sustainability performance. The 12-frame taxonomy provides a compact representation of certainty, responsibility, valence and legitimacy, while semantic triples and evidence links make report-level indicators inspectable. The six-case demonstration suggests that commitment-delivery imbalance, limited external-authority signals and selective positive narrative emphasis can be surfaced in a form relevant to sustainable-finance review.

The paper makes three principal contributions. First, it extends ESG analytics beyond topic and sentiment analysis by operationalising communication framing at claim and report levels. Second, it proposes an explainable AI architecture in which aggregate risk signals remain connected to identifiable textual evidence. Third, it demonstrates how narrative patterns can be operationalised as explainable screening signals for commitment-delivery imbalance, weak evidential support and disclosure-credibility gaps. Their defensible role is evidence triage and documented escalation, not factual, legal, assurance or investment conclusions.

Principal limitations include the small purposive corpus, selection and spectrum bias, no independent classifier validation, illustrative indicator weights and limited multilingual or adversarial testing. Future research should build a multi-company, multi-year corpus, establish expert-annotated gold standards, compare classification approaches and connect narrative indicators to assurance, controversies, target achievement and analyst decisions. Production deployment would also require versioned taxonomies, model documentation, data lineage, contestability and human oversight.

Data availability: The prototype dashboard used for the demonstration is available at <https://esg-fa.fly.dev/mfc> (accessed 9 August 2026).

REFERENCES

- Advertising Standards Authority. (2023, June 7). *ASA ruling on Shell UK Ltd t/a Shell*.
- Berg, F., Koelbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315-1344. <https://doi.org/10.1093/rof/rfac033>
- Conlon, S. J., Hale, J. G., Lukose, S., & Strong, J. (2008). Information extraction agents for service-oriented architecture using web service systems: A framework. *Journal of Computer Information Systems*, 48(3), 74-83. <https://doi.org/10.1080/08874417.2008.11646023>
- Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, 54(1), 64-87. <https://doi.org/10.1525/cmr.2011.54.1.64>
- Goffman E. (1974) *Frame Analysis: An Essay on the Organization of Experience*. Cambridge, MA: Harvard University Press.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- European Consumer Organisation. (2023). *Unbottling greenwashing: Lifting the lid on plastic bottle recycling claims*. BEUC.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75-105. <https://doi.org/10.2307/25148625>
- Higgins, C., & Walker, R. (2012). Ethos, logos, pathos: Strategies of persuasion in social/environmental reports. *Accounting Forum*, 36(3), 194-208. <https://doi.org/10.1016/j.accfor.2012.02.003>
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35-65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, norms, and selective disclosure: A global study of greenwashing. *Organization Science*, 27(2), 483-504. <https://doi.org/10.1287/orsc.2015.1039>
- Mitri, M. (2022). Story analysis using natural language processing and interactive dashboards. *Journal of Computer Information Systems*, 62(2), 216-226. <https://doi.org/10.1080/08874417.2020.1774442>
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45-77. <https://doi.org/10.2753/MIS0742-1222240302>
- U.S. Securities and Exchange Commission. (2023, September 25). *Deutsche Bank subsidiary DWS to pay \$25 million for anti-money laundering violations and misstatements regarding ESG investments* (Press Release No. 2023-194).