Indonesian Journal of Language and Economic Discourse
Vol. 1, No. 2, June 2026, pp. 84~97
ISSN: 3163-4362, DOI:

Green promises and unverifiable claims: Detecting corporate greenwashing in Indonesian sustainability reports through corpus linguistics and natural language processing

Arif Nugroho
Universitas Islam Negeri Raden Mas Said, Surakarta, Indonesia

Article Info

Article history:

Received Apr 29, 2026
Revised May 23, 2026
Accepted Jun 30, 2026

Keywords:

corporate greenwashing;
corpus linguistics;
documentary verifiability;
multimodal legitimation;
natural language processing.

ABSTRACT

Background: Corporate sustainability reports increasingly combine ambitious environmental commitments with uneven evidence, creating difficulties in distinguishing accountable disclosure from potentially misleading green claims. **Objective:** This study examines how Indonesian corporations construct environmental responsibility and whether linguistic promises, documentary specification, and visual legitimation remain mutually aligned. **Method:** A comparative corpus-assisted design integrates rule-based natural language processing, claim-verifiability assessment, and multimodal coding across six official sustainability reports, thirty-six contextual passages, and six report covers from three corporations during 2023–2024. **Results:** Promissory language dominated the textual corpus, whereas quantified performance statements occurred less frequently and future-oriented markers extended across most coded passages. Documentary analysis classified most units as having low verifiability, with time boundaries appearing more often than baselines, operational scopes, or assurance references. **Implication:** Cover analysis revealed recurrent combinations of nature imagery, community actors, green colour salience, and corporate logos, producing strong visual narratives of ecological and social responsibility. **Novelty:** This study introduces an integrated diagnostic framework that conceptualises greenwashing risk as misalignment among linguistic projection, evidential completeness, and multimodal legitimacy rather than as a property of isolated positive words, numerical claims, or environmental images and enables transparent comparison across companies, reporting years, claim types, evidential attributes, and visual strategies while preserving cautious interpretive boundaries in corporate reporting.

Corresponding Author:

Arif Nugroho
Department of English Education, Universitas Islam Negeri Raden Mas Said, Surakarta, Indonesia
Jl. Pandawa, Dusun IV, Pucangan, Kec. Kartasura, Sukoharjo, Jawa Tengah, 57168, Indonesia
Email: arif.nugroho@staff.uinsaid.ac.id

1. INTRODUCTION

Sustainability reporting has become an arena in which accountability and reputational management intersect. In Indonesia, Financial Services Authority Regulation No. 51/POJK.03/2017 requires financial institutions, issuers, and public companies to prepare sustainability reports, thereby expanding the volume of corporate environmental claims available to investors and citizens. Yet disclosure growth does not guarantee evidential quality: firms may foreground ambitious commitments while omitting baselines, operational boundaries, deadlines, or independent assurance. Internationally, Lagasio (2024) [1] analysed reports from 749 listed companies and demonstrated that discrepancies between portrayed and actual sustainability

practices can be investigated computationally, while Gorovaia and Makrominas (2024) [2] found that environmentally sanctioned firms produced longer, more positive, and less readable reports. Such evidence makes linguistic scrutiny socially consequential because unverifiable claims can influence investment judgments, corporate legitimacy, and public understandings of environmental performance [3], [4]. Indonesian reporting therefore requires analysis that separates environmental visibility from documentary accountability.

Previous scholarship has established several foundations for detecting greenwashing in corporate communication. Legitimacy-oriented research explains why firms strategically expand environmental disclosure when reputational or regulatory pressures intensify [5], while studies of impression management distinguish substantive reporting from symbolic presentation [6]. Corpus and NLP research has operationalised environmental vocabulary, sentiment, readability, and disclosure intensity to identify suspicious reporting patterns [2], [7]. More recent work has developed severity indices, transformer-based classifiers, explainable NLP systems, and machine-learning comparisons between discourse and external performance indicators [1], [8], [9]. However, existing approaches remain concentrated in European, Chinese, or cross-national datasets and frequently treat lexical positivity as a proxy for deception. Indonesian reports remain substantially underexamined, particularly through an integrated framework connecting linguistic construction, claim-level verifiability, and visual legitimation without equating textual risk signals with confirmed misconduct.

Addressing this gap, this study examines how Indonesian corporations linguistically and visually construct environmental responsibility and how far their claims remain documentarily verifiable. It asks three questions. First, which lexical, grammatical, and rhetorical patterns characterise green promises, including future-oriented commitments, vague environmental evaluation, quantification, and suppression of responsible actors? Second, to what extent do environmental statements specify measurable indicators, baselines, reporting scopes, deadlines, standards, supporting references, and independent assurance? Third, how do cover images, community actors, natural landscapes, colour palettes, corporate logos, and compositional arrangements reinforce or complicate textual sustainability narratives? Methodologically, this study combines corpus-assisted discourse analysis, rule-based and explainable NLP, a claim-verifiability matrix, and multimodal coding. This design follows calls for interpretable computational assessment rather than opaque classification [8] and responds to concerns that artificial intelligence can both strengthen sustainability assurance and facilitate increasingly polished greenwashing [10], [11].

This study advances the provisional argument that greenwashing risk is most visible not in isolated positive words but in discrepancies between environmental promise, evidential specification, and multimodal legitimation. Corporate claims become less accountable when ambitious futures are expressed without actors, measurable baselines, scopes, or traceable verification; conversely, quantification alone does not ensure credibility when metrics are selectively framed or disconnected from performance contexts. Visual narratives may naturalise corporate virtue by associating brands with landscapes, communities, and green colour schemes, even where accompanying claims remain partially verifiable. Accordingly, this study treats corpus and NLP indicators as diagnostic signals rather than proof of deception, consistent with research emphasising discourse–practice gaps and external validation [12], [9]. Its implication is methodological: credible greenwashing detection requires triangulating linguistic form, documentary evidence, and visual rhetoric, shifting analysis from counting sustainability language toward evaluating whether corporate promises can be independently examined.

2. LITERATURE REVIEW

2.1. Greenwashing

Greenwashing is understood as communication that presents environmental responsibility more favourably than underlying practices warrant, yet definitions differ in scope and evidential threshold. Some studies treat it as selective disclosure, whereby positive achievements are amplified and negative impacts suppressed [13], [14]. Others conceptualise it as a discourse–performance gap requiring comparison between claims and environmental outcomes [12], [9]. A third approach emphasises reputational manipulation through ambiguous, exaggerated, or unverifiable language [4], [15]. These differences matter because linguistic suspicion alone cannot establish deception without independent corroboration.

Research identifies greenwashing through interacting linguistic, organisational, and evidential indicators. Linguistic signals include vague environmental vocabulary, excessive positivity, promissory modality, passive constructions, and strategic omission of actors [2], [16]. Evidential indicators include absent baselines, undefined scopes, weak target specificity, selective quantification, and missing assurance [17], [18]. Organisational explanations emphasise legitimacy pressures, governance quality, reporting regulation, and materiality practices [5], [19], [20]. Consequently, greenwashing should be assessed as a layered reporting condition rather than a single feature.

2.2. Corpus linguistics and natural language processing

Corpus linguistics and natural language processing provide complementary approaches for examining sustainability discourse at scale. Corpus methods prioritise recurrent lexical patterns, collocations, concordances, stance markers, and semantic prosodies, enabling interpretation of how environmental responsibility is discursively normalised. NLP extends this capacity through automated classification, sentiment analysis, readability measurement, topic modelling, embeddings, and transformer-based prediction [21], [7]. Yet computational studies vary between descriptive text mining, supervised detection, and performance-linked auditing [1], [2]. This distinction is crucial because algorithmic classification may detect suspicious reporting patterns without demonstrating that a claim is false.

Recent detection frameworks organise textual evidence into several analytical dimensions. Lexical indicators measure environmental terminology, evaluative intensity, uncertainty, and future orientation; syntactic indicators capture passivisation, nominalisation, and agency suppression; discourse indicators examine thematic prominence, repetition, and selective emphasis [11], [22]. Machine-learning approaches add sentiment, readability, semantic similarity, contradiction detection, and document-level greenwashing scores [23], [24]. Explainable NLP further requires transparent features and interpretable rationales rather than opaque labels [8]. Accordingly, computational reliability depends on annotated definitions, validation procedures, and separation between prediction, interpretation, and verification.

2.3. Documentary verifiability

Documentary verifiability concerns whether environmental claims contain information for independent examination. Unlike truth, which requires external performance evidence, verifiability refers to whether a claim specifies measurable indicators, baselines, deadlines, boundaries, methodologies, and supporting sources. This distinction aligns with research on materiality, assurance, and disclosure quality, which treats transparent specification as a precondition for accountability rather than proof of performance [20], [25]. Other scholarship stresses that reporting can remain persuasive despite technical detail when metrics are selectively framed [26], [27]. Verifiability therefore mediates between linguistic form and empirical assessment.

Verifiability is strengthened by quantified metrics, explicit baselines, time-bound targets, defined scopes, named standards, traceable evidence, and independent assurance. It is weakened by aspirational wording, denominator-free percentages, shifting boundaries, selective timeframes, unsupported causal language, and persuasive sustainability symbolism. Multimodal research suggests that landscapes, communities, green colour palettes, and logos can legitimate environmental narratives beyond written claims [28], [10]. This study therefore theorises greenwashing risk as an alignment problem among linguistic construction, evidential specification, and visual legitimation. Its position is diagnostic: textual and visual signals identify risk, whereas confirmed greenwashing requires external performance comparison.

3. METHOD

This study analysed six publicly accessible sustainability reports issued by PT Astra International Tbk, PT Telkom Indonesia (Persero) Tbk, and PT Unilever Indonesia Tbk for reporting years 2023 and 2024. Reports represented diversified industry, telecommunications and digital infrastructure, and fast-moving consumer goods, enabling controlled cross-sectoral comparison within Indonesia. Primary material comprised complete official PDF reports and thirty-six context-preserving textual units selected from environmentally relevant passages. Supporting material included one reporting standard and one national regulatory document. Each document received a stable identifier, while every passage retained its page number, source link, title, date, language, quotation, surrounding context, and visual metadata.

This study adopted a comparative corpus-assisted mixed-method design integrating quantitative text profiling, qualitative discourse interpretation, documentary verification, and multimodal analysis. Corpus linguistics established recurring lexical and grammatical tendencies across reports, while natural language processing supported systematic feature extraction at passage level. A claim-verifiability matrix evaluated whether environmental statements contained measurable and traceable evidence, and multimodal coding examined how visual composition supported corporate legitimacy. This design did not infer deception from linguistic form alone. Instead, it treated promissory language, vagueness, selective quantification, weak specification, and persuasive imagery as diagnostic indicators requiring contextual interpretation and, where available, comparison with independent performance evidence.

Information was obtained exclusively from official corporate report portals and direct PDF repositories maintained by Astra International, Telkom Indonesia, and Unilever Indonesia. Regulatory context was drawn from Indonesia's Financial Services Authority, while disclosure criteria were informed by Global Reporting Initiative standards. Corporate reports constituted primary data because they contained

environmental claims, targets, metrics, images, and narrative explanations produced by reporting organisations. GRI material functioned as secondary analytical guidance, whereas OJK regulation supplied institutional context. For reproducibility, metadata recorded canonical URLs, access dates, publication information, reporting periods, language, sector, document type, inclusion basis, exclusion status, filename, provenance, and SHA-256 file hashes.

Table 1. Composition of dataset

Code	Data type	Source	Sector	Period	Quantity	Characteristics
IDN-ASII-SR-2023	Sustainability report	PT Astra International Tbk	Diversified industry	2023	1 report; 6 units	Indonesian–English; searchable PDF; visual cover data
IDN-ASII-SR-2024	Sustainability report	PT Astra International Tbk	Diversified industry	2024	1 report; 6 units	Indonesian–English; searchable PDF; visual cover data
IDN-TLKM-SR-2023	Sustainability report	PT Telkom Indonesia (Persero) Tbk	Telecommunications and digital infrastructure	2023	1 report; 6 units	Indonesian; searchable PDF; visual cover data
IDN-TLKM-SR-2024	Sustainability report	PT Telkom Indonesia (Persero) Tbk	Telecommunications and digital infrastructure	2024	1 report; 6 units	Indonesian; searchable PDF; visual cover data
IDN-UNVR-SR-2023	Sustainability report	PT Unilever Indonesia Tbk	Fast-moving consumer goods	2023	1 report; 6 units	Indonesian–English; searchable PDF; visual cover data
IDN-UNVR-SR-2024	Sustainability report	PT Unilever Indonesia Tbk	Fast-moving consumer goods	2024	1 report; 6 units	Indonesian–English; searchable PDF; visual cover data
GLOBAL-GRI-STANDARD-2021	Reporting standard	Global Reporting Initiative	Disclosure benchmark	2021	1 source	Used to assess disclosure completeness and evidential specification
IDN-OJK-REG-2017-051	Regulation	Otoritas Jasa Keuangan	Regulatory framework	2017	1 source	National sustainability-reporting context

Data collection proceeded through predefined Indonesian and English search strings covering sustainability reports, emissions, net zero, renewable energy, environmental targets, waste, water, and decarbonisation. Eligible documents were official-domain, publicly accessible, searchable sustainability reports concerning Indonesian operating entities during 2023–2024. News items, advertisements, third-party summaries, ordinary financial statements, inaccessible files, and duplicated editions were excluded. Retrieved PDFs were archived, hashed, converted to page-preserving text, segmented into contextual passages, and assigned document and unit codes. Whitespace-normalised lowercase hashes removed duplicate passages. Researchers then retained relevant quotations with surrounding context and recorded cover-level actors, symbols, colours, layouts, logos, frames, transcripts, and screenshot references.

Analysis followed three linked instruments. First, Corpus-Assisted Green Claim Profiling measured keywords, concordances, collocations, future markers, vague environmental lexemes, quantified expressions, passives, and agency suppression. Second, Claim Verifiability and Evidence Matrix coded metrics, baselines, deadlines, scopes, standards, assurance references, and a transparent zero-to-five evidence score. Third, Multimodal Legitimation Audit coded environmental imagery, community actors, colour salience, symbols, logo placement, layout, framing, and text–image relations. Automated coding was treated as provisional; all passages required manual validation, systematic cross-document comparison, and interpretive checking before reporting patterns or drawing conclusions about greenwashing risk.

4. RESULTS

4.1. Linguistic construction of green promises

Corporate environmental discourse was organised principally around prospective commitments rather than uniformly documented accounts of completed performance. Among thirty-six coded passages, twenty-five were classified as promissory language, six as quantified performance, and five as general green

positioning. Future or commitment markers appeared in twenty-nine passages, including several units whose principal classification was quantified performance or general positioning. This overlap indicates that measurable achievements were frequently connected to future aspirations, long-term targets, or continuing commitments. Quantified claims were unevenly distributed: none occurred in Astra's 2023 or Unilever's 2024 selected units, whereas Astra 2024 and Telkom 2023 each contained two. The resulting pattern differentiates the dominant temporal orientation of environmental discourse from the narrower distribution of measurable performance statements.

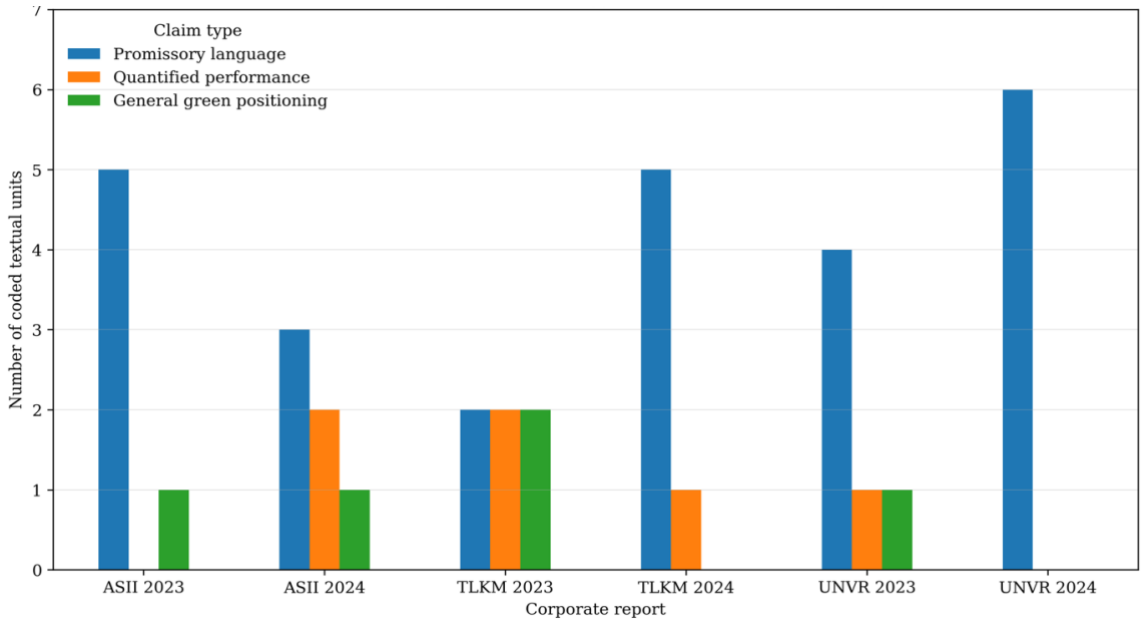

Figure 1. Distribution of linguistic claim types across sustainability reports

Promissory language accounted for 69.4% of coded units and appeared in every report, reaching all six passages in Unilever 2024 and five of six passages in Astra 2023 and Telkom 2024. By comparison, quantified performance represented only 16.7%, while general green positioning comprised 13.9%. Future-oriented expressions occurred in 80.6% of passages, yielding forty-four markers because several units combined commitment, target, and aspiration terms. Vague green lexemes appeared in thirteen passages, or 36.1%, and were most frequent in Astra 2024 and Unilever 2023. No passage met the specified rule for actor suppression. Objectively, therefore, this corpus contains a broad concentration of prospective and evaluative wording, accompanied by a considerably smaller set of passages reporting directly measurable environmental outcomes.

This distribution supports an understanding of sustainability reporting as a temporally managed legitimacy discourse in which future commitments extend corporate accountability into periods not yet open to verification. Strategic-language research similarly shows that environmental communication may combine favourable evaluation, aspiration, and selective specification to construct organisational responsibility [16], [2]. However, promissory wording cannot itself be treated as evidence of deception, particularly when a commitment includes deadlines, scopes, or measurable indicators. Its analytical significance arises from its dominance relative to quantified performance and from its co-occurrence with vague environmental vocabulary. Consistent with Montgomery et al. (2023) and Xu et al. (2023), this pattern positions greenwashing risk within the relationship between rhetorical ambition and evidential support, rather than within individual environmental terms considered in isolation.

Table 2. Linguistic construction of environmental claims in corporate sustainability reports

Subcategory or indicator	Operational criterion	Frequency	Percentage	Distribution across reports	Excerpts	Data code
Promissory language	Passage foregrounds commitments, aspirations, targets, intended actions, or future environmental improvement	25 units	69.4%	ASII 2023: 5; ASII 2024: 3; TLKM 2023: 2; TLKM 2024: 5; UNVR 2023: 4; UNVR 2024: 6	"with its commitment to reduce water intake intensity"	IDN-ASII-SR-2023-U002
Quantified performance	Passage includes a measurable environmental outcome expressed through percentages, volumes, energy units, emissions, area, or comparable metrics	6 units	16.7%	ASII 2024: 2; TLKM 2023: 2; TLKM 2024: 1; UNVR 2023: 1	"pemanfaatan energi terbarukan sebesar 194.093 MWh"	IDN-ASII-SR-2024-U002
General green positioning	Passage associates corporate activities with sustainability, environmental management, or decarbonisation without presenting a complete measurable claim	5 units	13.9%	ASII 2023: 1; ASII 2024: 1; TLKM 2023: 2; UNVR 2023: 1	"strategi dekarbonisasi, dan carbon offsetting"	IDN-ASII-SR-2023-U001
Future and commitment markers	Passage contains markers such as akan, berkomitmen, menargetkan, menuju, aim, will, commit, or target	29 units	80.6%	ASII 2023: 5; ASII 2024: 4; TLKM 2023: 3; TLKM 2024: 6; UNVR 2023: 5; UNVR 2024: 6	"Pengurangan limbah mencapai 5 ton pada 2030"	IDN-TLKM-SR-2023-U005
Vague green lexemes	Passage uses broad positive terms such as sustainable, green, environmentally friendly, sustainable growth, positive impact, or similar evaluative wording	13 units	36.1%	ASII 2023: 2; ASII 2024: 4; TLKM 2024: 1; UNVR 2023: 4; UNVR 2024: 2	"Astra recognizes the need to decarbonize our operations"	IDN-ASII-SR-2023-U003
Actor suppression	Passive or agentless wording conceals the party responsible for implementing, measuring, or verifying environmental action	0 units	0.0%	No coded occurrence	—	—
Total future markers	Aggregate occurrence of future-oriented and commitment markers across all coded passages	44 markers	—	Distributed across 29 units	"commitment," "target," "menuju," and comparable forms	Multiple units
Total vague green lexemes	Aggregate occurrence of selected broadly positive environmental expressions	18 markers	—	Distributed across 13 units	"berkelanjutan," "lestari," "sustainable," and "green"	Multiple units

4.2. From environmental claims to documentary evidence

Environmental claims displayed markedly weaker documentary specification than their strong promissory orientation might suggest. Using the same thirty-six passages analysed previously, twenty-two units were classified as having low verifiability and fourteen as having partial verifiability; no unit reached the higher-verifiability threshold. Classification was based on five observable elements: measurable metrics, explicit baselines, temporal boundaries, defined scopes or reporting standards, and assurance references. This distribution does not establish that the claims were inaccurate. Rather, it indicates that most retained passages lacked enough internally available information to permit systematic examination without consulting other sections, appendices, or external evidence. Documentary accountability was therefore unevenly embedded within corporate environmental language, with limited convergence among quantification, temporal specificity, operational boundaries, comparative reference points, and independent validation.

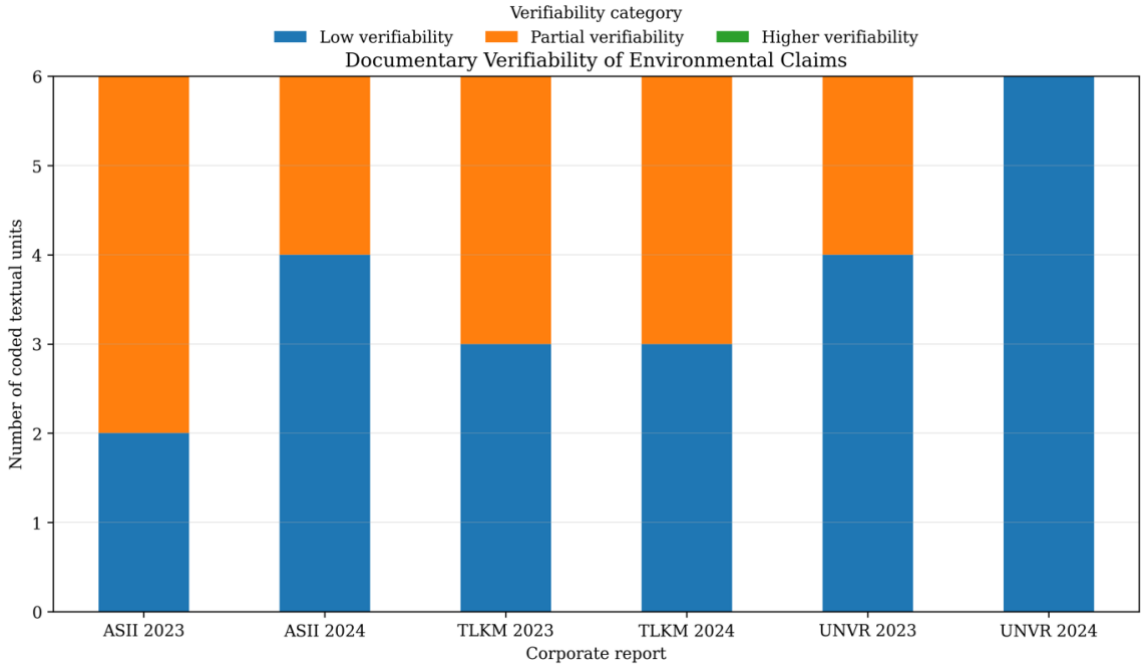

Figure 2. Documentary verifiability of environmental claims

Low-verifiability units represented 61.1% of the corpus, while partial-verifiability units accounted for 38.9%. Unilever 2024 contained six low-verifiability passages and no partially verifiable unit, whereas Astra 2023 contained four partially verifiable and two low-verifiability passages. Temporal references were the most common evidential feature, occurring in twelve units, followed by measurable metrics in six units. Only one passage specified an operational boundary or reporting scope, and no coded unit contained an explicit baseline or assurance reference. The average composite evidence score was 0.53 on a five-point scale. These distributions show that deadlines and reporting years were more common than comparative baselines, scoped methodologies, or external verification, producing claims that were temporally located but often insufficiently anchored for longitudinal or independent assessment.

Theoretically, this pattern distinguishes disclosure visibility from evidential accountability. Materiality and disclosure research argues that corporate reports become more decision-useful when environmental statements specify comparable indicators, reporting boundaries, and traceable foundations rather than relying primarily on narrative prominence [20], [25]. Likewise, studies of climate disclosure caution that numerical expressions can create an appearance of precision while remaining difficult to assess when baselines, denominators, scopes, or assurance mechanisms are absent [17], [18]. In this corpus, temporal targets frequently strengthened the appearance of commitment, yet rarely converged with the full evidential architecture required for robust verification. Greenwashing risk therefore emerges from incomplete alignment between promise, measurement, comparison, and validation, not from aspiration or quantification considered separately.

Table 3. Documentary verifiability of environmental claims

Evidential subcategory	Operational indicator	Frequency	Percentage	Distribution across reports	Excerpts	Data code
Low verifiability	No coded metric, baseline, time boundary, scope or reporting standard, or assurance reference	22	61.1%	ASII 2023: 2; ASII 2024: 4; TLKM 2023: 3; TLKM 2024: 3; UNVR 2023: 4; UNVR 2024: 6	"with its commitment to reduce water intake intensity"	IDN-ASII-SR-2023-U002
Partial verifiability	Passage contains one or two evidential elements but does not reach the threshold for higher verifiability	14	38.9%	ASII 2023: 4; ASII 2024: 2; TLKM 2023: 3; TLKM 2024: 3; UNVR 2023: 2; UNVR 2024: 0	"Pengurangan limbah mencapai 5 ton pada 2030"	IDN-TLKM-SR-2023-U005
Higher verifiability	At least three of five evidential elements are present within the same coded unit	0	0.0%	No report contained a qualifying unit	—	—
Metric present	Percentage, emissions quantity, energy unit, waste amount, area, or another measurable expression	6	16.7%	ASII 2024: 2; TLKM 2023: 2; TLKM 2024: 1; UNVR 2023: 1	"pemanfaatan energi terbarukan sebesar 194.093 MWh"	IDN-ASII-SR-2024-U002
Time boundary present	Reporting year, target year, or explicit temporal reference occurs in the passage	12	33.3%	ASII 2023: 4; ASII 2024: 2; TLKM 2023: 2; TLKM 2024: 3; UNVR 2023: 1; UNVR 2024: 0	"target kuantitatif tahun 2030"	IDN-TLKM-SR-2024-U003
Baseline present	Explicit comparison year, initial measurement, or baseline value appears in the coded passage	0	0.0%	No coded occurrence	—	—
Scope or standard present	Operational scope, emissions scope, reporting boundary, or named disclosure standard is specified	1	2.8%	TLKM 2024: 1	"penurunan emisi GRK cakupan 1 dan 2 di tahun 2030"	IDN-TLKM-SR-2024-U005
Assurance reference	Independent assurance, audit, verification, or equivalent validation is named in the passage	0	0.0%	No coded occurrence	—	—
Mean evidence score	Additive score for metric, baseline, time boundary, scope or standard, and assurance	0.53 of 5	10.6% of maximum	Based on all 36 coded units	—	Multiple units

4.3. Visualising Sustainability and Corporate Legitimacy

Visual legitimation was examined at the document-cover level while retaining direct links to the same six reports and thirty-six textual units used in preceding analyses. Each report contributed one cover record coded for nature imagery, community actors, green-colour salience, and visible corporate identity, alongside descriptive fields for symbols, layout, framing, logo position, and transcribed cover text. Twenty-one of twenty-four possible binary features were present, producing an overall visual-feature density of 87.5%. Five covers combined all four indicators, whereas Astra 2023 contained only a salient corporate logo. The covers therefore formed two broad representational orientations: image-rich compositions connecting sustainability with nature and society, and a minimalist institutional design relying on typography, white space, and corporate identity.

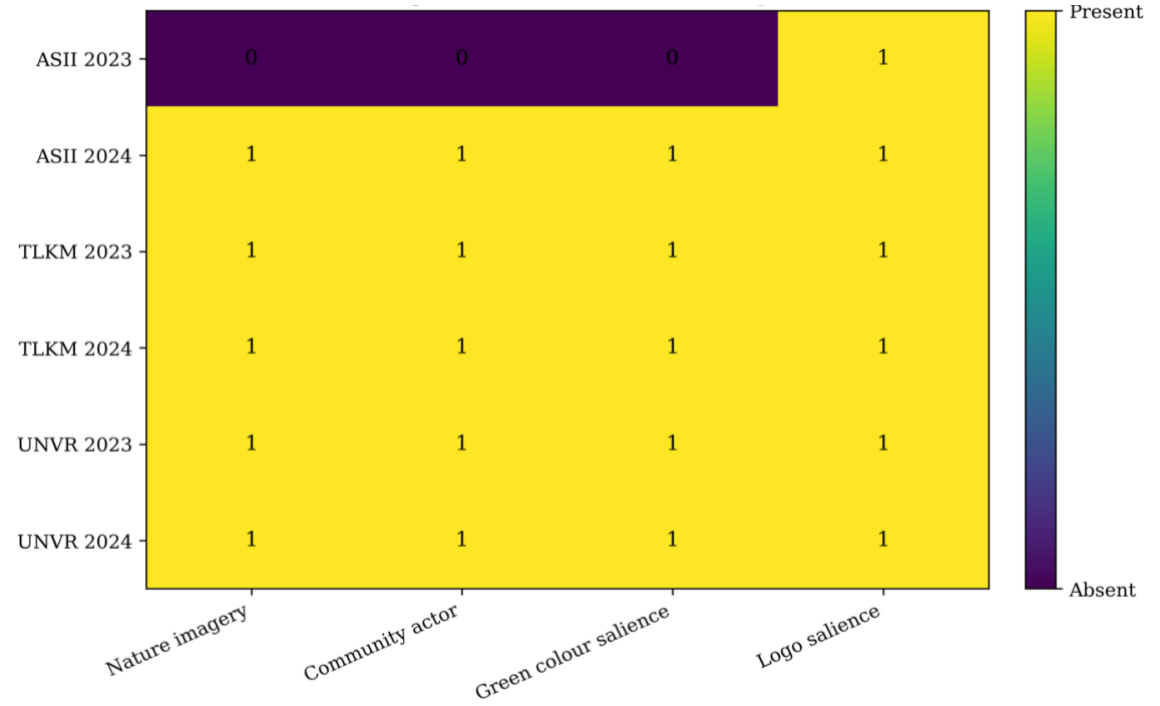

Figure 3. Multimodal legitimation features across report covers

Nature imagery, community representation, and green-colour salience each occurred on five of six covers, while corporate logos appeared on all six. Telkom's two reports used full-bleed photographs that situated social actors within rural or plantation landscapes. Astra 2024 and Unilever 2023 employed illustrations combining community activity, agriculture, housing, mobility, commerce, and environmental technology. Unilever 2024 used a photographic collage linking agricultural production, consumers, commercial facilities, and family life. Astra 2023 differed through its sparse white composition, bilingual report title, turquoise-blue accents, and absence of human or ecological imagery. Objectively, visual consistency was strongest in logo presence and in the recurring combination of green settings with socially identifiable beneficiaries across five reports.

These patterns indicate that sustainability is visually legitimated by associating corporate identity with ecological abundance, social inclusion, and recognisable public benefit. Multimodal legitimacy does not depend solely on explicit environmental symbols; it also emerges through actors, settings, spatial composition, and colour, which organise preferred interpretations before detailed disclosure is read. This observation aligns with research showing that sustainability communication can extend persuasive environmental positioning beyond propositional claims into images, branding, and platform-specific presentation [28], [10]. Importantly, visual density did not itself establish greenwashing. Its analytical relevance lies in its relationship with preceding findings: five visually saturated covers accompanied a corpus dominated by promissory language and limited documentary verifiability, creating a potential asymmetry between representational confidence and evidential completeness.

Table 4. Multimodal legitimation features across corporate sustainability report covers

Visual dimension	Operational indicator	Frequency	Percentage	Distribution across reports	Visual manifestation	Document codes
Nature imagery	Presence of landscapes, agriculture, plantations, fields, renewable-energy symbols, or other environmental scenery	5	83.3%	ASII 2024; TLKM 2023; TLKM 2024; UNVR 2023; UNVR 2024	Agricultural landscapes, plantations, green fields, solar panels, wind turbines, and nature-oriented settings	IDN-ASII-SR-2024; IDN-TLKM-SR-2023; IDN-TLKM-SR-2024; IDN-UNVR-SR-2023; IDN-UNVR-SR-2024
Community actors	Presence of workers, farmers, consumers, children, families, or other community members	5	83.3%	ASII 2024; TLKM 2023; TLKM 2024; UNVR 2023; UNVR 2024	Illustrated workers, rural residents, schoolchildren, farmers, consumers, and family members	Same five document codes
Green colour salience	Green functions as a dominant or prominent visual colour	5	83.3%	ASII 2024; TLKM 2023; TLKM 2024; UNVR 2023; UNVR 2024	Green-dominant illustrations, fields, plantations, and turquoise-green corporate compositions	Same five document codes
Corporate logo salience	Corporate identity appears visibly on the cover	6	100.0%	All reports	Logos positioned in upper or lower cover areas and integrated with report titles	All six document codes
Minimalist corporate framing	Cover relies primarily on typography, white space, report title, and logo rather than environmental or social imagery	1	16.7%	ASII 2023	White background, bilingual title, turquoise-blue accents, no human actor or natural landscape	IDN-ASII-SR-2023
Photographic ecological framing	Full-bleed photography places social actors within natural or productive landscapes	2	33.3%	TLKM 2023; TLKM 2024	Rural shelter, fields, plantation, sunlight, adult educator, and schoolchildren	IDN-TLKM-SR-2023; IDN-TLKM-SR-2024
Illustrated systemic framing	Illustration integrates environmental, economic, technological, and community symbols within one composition	2	33.3%	ASII 2024; UNVR 2023	Homes, mobility, agriculture, renewable energy, commerce, and community activity	IDN-ASII-SR-2024; IDN-UNVR-SR-2023
Photographic collage framing	Multiple photographs combine production, consumption, community, and family settings	1	16.7%	UNVR 2024	Agriculture, young consumers, retail facilities, and parent–child interaction	IDN-UNVR-SR-2024
Visual feature density	Total presence of four coded binary features across six covers	21 of 24 possible features	87.5%	Five covers contain all four features; ASII 2023 contains logo salience only	Combined nature, community, green colour, and logo indicators	All six document codes

5. DISCUSSION

The dominance of promissory language over quantified performance indicates that corporate sustainability reporting functions primarily as a temporal architecture of legitimacy. Its communicative function is to project responsibility forward, allowing firms to appear aligned with environmental expectations before commitments are fully measurable. This does not render future-oriented discourse inherently deceptive; long-term transition strategies necessarily depend on targets and aspirations. Dysfunction emerges when promises become the principal carrier of environmental credibility while evidential specification remains secondary. This pattern supports Kochkina et al. (2024), who show that strategic language shapes sustainability narratives, and Gorovaia and Makrominas (2024) [2], who identify positivity and textual expansion as possible reporting signals. It also complicates studies that treat quantified disclosure as sufficient evidence of substantive commitment. In this study, rhetorical ambition appears analytically significant because it structures stakeholder expectations, yet its legitimacy depends on whether subsequent reporting converts projected intent into traceable, comparable, and accountable environmental performance.

The prevalence of future-oriented commitments can be explained by structural pressures within sustainability reporting rather than by lexical preference alone. Corporations operate within regulatory, reputational, and investor-facing environments that reward alignment with environmental agendas, even when transitions remain incomplete. Legitimacy theory suggests that firms respond to such pressures by communicating conformity with valued norms [5], while materiality research shows that organisations retain discretion over which issues, indicators, and time horizons receive emphasis [20]. Promissory discourse therefore offers flexibility: it signals responsiveness, defers evaluation, and reduces the immediate burden of demonstrating completed transformation. This interpretation is consistent with Montgomery et al. (2023), who conceptualise greenwashing as a field of selective communication, but it contrasts with approaches that equate any ambitious target with manipulation. The underlying structure is better understood as an accountability delay, in which future language extends reputational benefit beyond presently verifiable environmental performance.

The predominance of low and partial verifiability has consequences for how sustainability reports can be used by investors, regulators, researchers, and civil society. Reports may remain broadly informative as statements of orientation, but their decision usefulness declines when metrics are detached from baselines, operational boundaries, assurance, or comparable timeframes. The principal dysfunction is not necessarily falsehood; it is restricted meaningful examinability. Stakeholders cannot determine whether improvement is substantial, selective, or merely reformulated across reporting periods consistently. This finding supports Adams and Abhayawansa (2021) [25], who connect harmonised reporting with comparability, and Wedari et al. (2021), who question whether climate disclosures can conceal greenwashing risks. However, it also qualifies text-based detection approaches: a low evidence score should not be interpreted as definitive proof that environmental performance is poor. In this study, verifiability operates as an intermediate accountability condition, distinguishing persuasive disclosure from claims that can be checked independently, compared, and critically challenged.

Weak documentary verifiability is produced by the interaction of reporting discretion, fragmented standards, and organisational control over evidential boundaries. Companies determine where metrics appear, how scopes are defined, which baselines are selected, and whether assurance is attached to specific claims or only to reporting processes. Caputo et al. (2021) argue that regulation can improve transparency, yet reporting requirements do not automatically eliminate selective disclosure. De Silva Lokuwaduge and De Silva (2022) [18] similarly show that ESG disclosure may coexist with greenwashing risk when risk information lacks sufficient depth. The absence of combined evidential elements in single passages may reflect document architecture, where targets, methodologies, and assurance statements are dispersed across sections. This possibility prevents claims about intentional obscuration. Nevertheless, this structural effect remains important: fragmentation increases verification costs and privileges expert readers. Thus, incomplete evidence appears not merely as missing information, but as an arrangement that redistributes the burden of proof.

The visual concentration of nature, community actors, green colour, and corporate logos shows that sustainability legitimacy is constructed before readers encounter detailed textual evidence. These images function by translating abstract environmental commitments into recognisable scenes of ecological harmony, social benefit, and corporate stewardship. Their communicative value lies in accessibility: visual narratives can make sustainability agendas intelligible and emotionally resonant. Dysfunction arises when representational confidence exceeds documentary support, because images may stabilise favourable interpretations that textual claims cannot fully substantiate. This finding extends Oppong-Tawiah and

Webster's (2023) argument that greenwashing can operate through platformed corporate communication and complements De Villiers et al. (2023), who warn that advanced communication technologies may intensify polished sustainability narratives. Yet visual symbolism should not be reduced to manipulation. In this study, its significance derives from alignment: imagery becomes analytically problematic when ecological abundance and social inclusion are not matched by equally visible evidence, scope, and accountability.

The recurrence of visual elements across sectors suggests an underlying grammar of sustainability rather than isolated choices. Green landscapes naturalise environmental responsibility, community actors humanise corporate action, and logos bind these values to identity. These combinations reflect institutionalised conventions through which sustainability becomes recognisable and useful. Companies may adopt these conventions because they offer low-cost symbolic coherence across environmental issues, from emissions to waste and water. This explanation is compatible with legitimacy-oriented accounts of disclosure and with multimodal perspectives that treat meaning as distributed across language, image, colour, and composition. It differs from interpretations that presume visual saturation necessarily conceals poor performance. The evidence supports a conditional claim: visual legitimacy intensifies the persuasive force of reporting, but greenwashing risk emerges only when multimodal promise, linguistic commitment, and documentary evidence become misaligned. This study therefore positions greenwashing as a cross-modal accountability problem, not a property of words, numbers, or images independently.

6. CONCLUSION

This study demonstrates that corporate greenwashing risk is best understood as a problem of alignment among linguistic promise, documentary evidence, and visual legitimation. Indonesian sustainability reports were dominated by future-oriented commitments, while quantified performance claims remained less frequent and evidential elements were unevenly combined. Nature imagery, community actors, green colour, and corporate logos further strengthened the persuasive presentation of environmental responsibility. The principal lesson is that neither positive language, numerical disclosure, nor ecological imagery is independently sufficient to establish misleading communication. This study contributes an integrated corpus-assisted framework that connects discourse profiling, claim-level verifiability, and multimodal analysis without collapsing diagnostic indicators into proof of deception.

Several limitations delimit these conclusions. The corpus covered six reports from three companies and two reporting years, while textual selection focused on thirty-six environmentally relevant passages and cover-level visuals. Findings therefore cannot represent Indonesian corporate reporting generally, and passage-level coding may not capture evidence dispersed across appendices or technical sections. Initial rule-based NLP also requires broader human validation, intercoder assessment, and testing against independently verified environmental performance. Future research should expand sectoral and temporal coverage, analyse complete reports, incorporate corporate emissions and regulatory records, compare Indonesian and English versions, and develop supervised, explainable models trained on expert-annotated claims. Such triangulation would support stronger distinctions among aspiration, weak verifiability, and demonstrable greenwashing.

ACKNOWLEDGMENTS

The author thanks the reviewers and editorial team for their constructive feedback on this manuscript.

FUNDING INFORMATION

Author states no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Arif Nugroho: conceptualization, methodology, formal analysis, writing – original draft, writing – review and editing.

CONFLICT OF INTEREST STATEMENT

Author states no conflict of interest.

AI USE DECLARATION

The author used artificial intelligence tools to assist with the generation of illustrative figures and with the formatting and layout of this manuscript. AI was not used to generate the substantive content, conceptualization, data, analysis, or conclusions of the research, all of which remain the sole responsibility of the author.

INFORMED CONSENT

Not applicable – this study did not involve direct human participants. Data consisted solely of publicly accessible corporate sustainability reports and related institutional materials.

ETHICAL APPROVAL

This research did not involve human or animal subjects; it analyzed publicly accessible corporate sustainability reports and related institutional materials, and therefore no institutional review board approval was required. Where research does involve human subjects, the author affirms that such research would be conducted in full compliance with all relevant national regulations and institutional policies, in accordance with the tenets of the Declaration of Helsinki, and would be approved by the author's institutional review board or equivalent committee.

DATA AVAILABILITY

The corpus of documents analyzed in this study are publicly accessible and available from the corresponding author upon reasonable request.

REFERENCES

[1] V. Lagasio, "ESG-washing detection in corporate sustainability reports," *International Review of Financial Analysis*, Art. no. 103742, 2024, doi: 10.1016/j.irfa.2024.103742.

[2] N. Gorovaia and M. Makrominas, "Identifying greenwashing in corporate-social responsibility reports using natural-language processing," *European Financial Management*, vol. 31, pp. 427–462, 2025, doi: 10.1111/eufm.12509.

[3] T. Li, X. Shu, and G. Liao, "Does corporate greenwashing affect investors' decisions?," *Finance Research Letters*, Art. no. 105877, 2024, doi: 10.1016/j.frl.2024.105877.

[4] J. Keilmann and T. Koch, "When environmental claims are empty promises: How greenwashing affects corporate reputation and credibility," *Environmental Communication*, vol. 18, pp. 266–284, 2024, doi: 10.1080/17524032.2023.2267782.

[5] F. Akhter, M. R. Hossain, H. Elrehail, S. U. Rehman, and B. Almansour, "Environmental disclosures and corporate attributes, from the lens of legitimacy theory: A longitudinal analysis on a developing country," *European Journal of Management and Business Economics*, 2022, doi: 10.1108/EJMBE-01-2021-0008.

[6] Sneiderienе and R. Legenzova, "Uncovering greenwashing: Investigating impression management gap in corporate reporting," *Sustainability*, vol. 17, no. 18, Art. no. 8342, 2025, doi: 10.3390/su17188342.

[7] M. Gutierrez-Bustamante and L. Espinosa-Leal, "Natural language processing methods for scoring sustainability reports—A study of Nordic listed companies," *Sustainability*, vol. 14, no. 15, Art. no. 9165, 2022, doi: 10.3390/su14159165.

[8] K. Ong, R. Mao, R. Satapathy, R. Shirota Filho, E. Cambria, J. Sulaeman, and G. Mengaldo, "Explainable natural language processing for corporate sustainability analysis," *Information Fusion*, Art. no. 102726, 2024, doi: 10.1016/j.inffus.2024.102726.

[9] I. A. A. D. Santos, A. C. De Araújo Moxotó, and E. De Souza Mota, "Greenwashing in Brazilian corporations: A machine learning approach to unmask the discourse–practice gap," *Corporate Social Responsibility and Environmental Management*, 2025, doi: 10.1002/csr.70133.

[10] De Villiers, R. Dimes, and M. Molinari, "How will AI text generation and processing impact sustainability reporting? Critical analysis, a conceptual framework and avenues for future research," *Sustainability Accounting, Management and Policy Journal*, 2023, doi: 10.1108/SAMPJ-02-2023-0097.

[11] W. Moodaley and A. Telukdarie, "Greenwashing, sustainability reporting, and artificial intelligence: A systematic literature review," *Sustainability*, vol. 15, no. 2, Art. no. 1481, 2023, doi: 10.3390/su15021481.

[12] W. Xu, M. Li, and S. Xu, "Unveiling the "veil" of information disclosure: Sustainability reporting "greenwashing" and "shared value."," *PLOS ONE*, vol. 18, Art. no. e0279904, 2023, doi: 10.1371/journal.pone.0279904.

[13] R. Zharfpeykan, "Representative account or greenwashing? Voluntary sustainability reports in Australia's mining/metals and financial services industries," *Business Strategy and the Environment*, 2021, doi: 10.1002/bse.2744.

[14] S. Ruiz-Blanco, S. Romero, and B. Fernandez-Feijoo, "Green, blue or black, but washing—What company characteristics determine greenwashing?," *Environment, Development and Sustainability*, vol. 24, pp. 4024–4045, 2022, doi: 10.1007/s10668-021-01602-x.

[15] A. W. Montgomery, T. P. Lyon, and J. Barg, "No end in sight? A greenwash review and research agenda," *Organization & Environment*, vol. 37, pp. 221–256, 2024, doi: 10.1177/10860266231168905.

[16] N. Kochkina, S. Macchia, and M. Floris, "Strategic language use in sustainability reporting: An empirical study," *Sustainability*, vol. 16, no. 23, Art. no. 10229, 2024, doi: 10.3390/su162310229.

[17] L. K. Wedari, C. Jubb, and A. Moradi-Motlagh, "Corporate climate-related voluntary disclosures: Does potential greenwash exist among Australian high emitters' reports?," *Business Strategy and the Environment*, 2021, doi: 10.1002/bse.2836.

[18] S. De Silva Lokuwaduge and K. M. De Silva, "ESG risk disclosure and the risk of green washing," *Australasian Business, Accounting and Finance Journal*, vol. 16, no. 1, 2022, doi: 10.14453/aabfj.v16i1.10.

[19] F. Caputo, S. Pizzi, L. Ligorio, and R. Leopizzi, "Enhancing environmental information transparency through corporate social responsibility reporting regulation," *Business Strategy and the Environment*, 2021, doi: 10.1002/bse.2814.

[20] M. Farooq, R. Zaman, D. Sarraj, and F. Khalid, "Examining the extent of and drivers for materiality assessment disclosures in sustainability reports," *Sustainability Accounting, Management and Policy Journal*, 2021, doi: 10.1108/SAMPJ-04-2020-0113.

[21] H. Kang and J. Kim, "Analyzing and visualizing text information in corporate sustainability reports using natural language processing methods," *Applied Sciences*, vol. 12, no. 11, Art. no. 5614, 2022, doi: 10.3390/app12115614.

[22] A. Vinella, M. Capetz, R. Pattichis, C. Chance, and R. Ghosh, "Leveraging language models to detect greenwashing," *arXiv*, 2023, doi: 10.48550/arXiv.2311.01469.

[23] X. Wang, X. Gao, and M. Sun, "Construction and analysis of corporate greenwashing index: A deep learning approach," *EPJ Data Science*, vol. 14, 2025, doi: 10.1140/epjds/s13688-025-00562-w.

[24] A. Ahmad, S. Raza, and R., "Greenwashing in corporate climate disclosures: A machine learning-based detection approach," *Global Social Sciences Review*, 2025, doi: 10.31703/gssr.2025(x-iii).10.

[25] A. Adams and S. Abhayawansa, "Connecting the COVID-19 pandemic, environmental, social and governance investing and calls for harmonisation of sustainability reporting," *Critical Perspectives on Accounting*, vol. 82, Art. no. 102309, 2021, doi: 10.1016/j.cpa.2021.102309.

[26] M. Cerciello, F. Busato, and S. Taddeo, "The effect of sustainable business practices on profitability: Accounting for strategic disclosure," *Corporate Social Responsibility and Environmental Management*, 2022, doi: 10.1002/csr.2389.

[27] S. Ratti, M. Arena, G. Azzone, and L. Dell'Agostino, "Beyond corporate environmental disclosure: Tangible efforts or greenwashing? The case of Italian non-financial listed companies," *Sustainable Development*, 2025, doi: 10.1002/sd.3544.
[28] Oppong-Tawiah and J. Webster, "Corporate sustainability communication as "fake news": Firms' greenwashing on Twitter," *Sustainability*, vol. 15, no. 8, Art. no. 6683, 2023, doi: 10.3390/su15086683.

BIOGRAPHY OF AUTHOR

Arif Nugroho is a lecturer at the English Education Department, Universitas Islam Negeri Raden Mas Said Surakarta, Indonesia. His academic studies are related to English education, applied linguistics, and related interdisciplinary areas. His research includes computational linguistics, code-mixing, machine learning in language analysis, corpus linguistics, digital literacy, and technology-enhanced language learning. He can be contacted at email: arif.nugroho@staff.uinsaid.ac.id