

Climate risk and financial stability research: a computational bibliometric analysis of thematic framing (2010–2025)

Imad EL KHLIFI, (Ph.D. Student)

*Faculty of Economics and Management, Tangier
 Abdelmalek Essaâdi University (UAE), Tétouan, Morocco*

Mohamed DRISSI BAKHKHAT, (Full Professor)

*Faculty of Economics and Management, Tangier
 Abdelmalek Essaâdi University (UAE), Tétouan, Morocco*

Correspondence address:	Faculty of Economics and Management, Abdelmalek Essaâdi University (UAE), Tétouan, Morocco
Disclosure Statement:	The authors declare that they have not received any financial support that could have influenced the objectivity of this study. They take full responsibility for any potential plagiarism, the use of artificial intelligence in the writing process, as well as for the results presented in this article.
Conflict of Interest:	The authors report no conflicts of interest.
Cite this article:	EL KHLIFI, I., & DRISSI BAKHKHAT, M. (2026). Climate risk and financial stability research: a computational bibliometric analysis of thematic framing (2010–2025). <i>International Journal of Accounting, Finance, Auditing, Management and Economics</i> , 7(13), 168–193. https://doi.org/10.5281/zenodo.23171775
License	This is an open access article under the CC BY-NC-ND license

Received: 14/09/2026

Accepted: 14/10/2026

Climate risk and financial stability research: a computational bibliometric analysis of thematic framing (2010–2025)

Abstract:

Research on climate risk and financial stability has grown rapidly, yet few studies jointly address its bibliometric structure, thematic framing (risk versus solution), and underlying data reliability. To address this gap, this study analyzes over 4,100 relevant English-language articles from the Dimensions database (2010–2025). By combining a domain lexicon, a manual audit, and machine learning models (SVM, ClimateBERT, FinBERT), we classify the abstracts' thematic orientation. Our manual audit reveals substantial noise in the raw data, highlighting the necessity of strict relevance filtering. Methodologically, while lexicon-based classification shows weak agreement with human coding, machine learning models deliver robust performance. Empirically, we find no clear temporal trend in the share of solution-oriented research. However, regression models demonstrate that solution-oriented papers receive a significant citation premium (approximately 27% higher expected citations), an association that remains robust across stricter filters and journal fixed effects. These findings emphasize the crucial role of data curation in computational bibliometrics and suggest that solution-focused research attracts greater academic attention.

Keywords: Climate risk, financial stability, bibliometric analysis, Dimensions, thematic framing, text classification, machine learning.

JEL Classification : C38, C45, G10, G18, Q54

Paper type : Empirical quantitative research (computational bibliometrics)

Résumé :

La recherche sur le risque climatique et la stabilité financière s'est développée rapidement, mais peu d'études examinent conjointement sa structure bibliométrique, son cadrage thématique (risque versus solution) et la fiabilité des données sous-jacentes. Pour combler cette lacune, cette étude analyse plus de 4 100 articles en anglais issus de la base Dimensions (2010-2025). En combinant un lexique de domaine, un audit manuel et des modèles d'apprentissage automatique (SVM, ClimateBERT, FinBERT), nous classons l'orientation thématique des résumés. Notre audit révèle un bruit important dans les données brutes, soulignant la nécessité d'un filtrage rigoureux. Sur le plan méthodologique, si le lexique montre une faible concordance avec le codage humain, les modèles d'apprentissage automatique offrent des performances robustes. Empiriquement, la part des recherches orientées solutions ne présente pas de tendance temporelle claire. Toutefois, les modèles de régression montrent que ces articles bénéficient d'une prime de citation significative (environ 27 % de citations attendues en plus), une association qui résiste aux effets fixes de revue et à des filtres stricts. Ces résultats soulignent l'importance de la curation des données en bibliométrie computationnelle et suggèrent que la recherche axée sur les solutions suscite davantage d'attention académique.

Mots clés : Risque climatique, stabilité financière, analyse bibliométrique, Dimensions, cadrage thématique, classification de textes, apprentissage automatique.

Classification JEL : C38, C45, G10, G18, Q54

Type du papier : Recherche empirique quantitative (bibliométrie computationnelle)

1. Introduction

Climate change is increasingly treated as a source of financial risk. Central banks, supervisors and investors have moved from describing it as an environmental externality to analysing it as a determinant of financial stability (Carney, 2015; NGFS, 2019; IPCC, 2021). Physical damages, the repricing of carbon-intensive assets and liability exposures can propagate through markets, banks and insurers (Battiston et al., 2017; Campiglio et al., 2018; Acharya et al., 2023), and foundational work on the economics of climate change and of catastrophic risk explains why these risks are difficult to price (Stern, 2008; Weitzman, 2009).

Scholarly production on this topic has followed the policy agenda, and the resulting body of work is too large to be surveyed narratively. Bibliometric methods are a standard way to describe such bodies of work (Zupic and Čater, 2015; Donthu et al., 2021), and they have been applied to sustainability research (Leal Filho et al., 2018). Two features of this literature make a closer look useful. First, it is contested, in particular on whether transition risk is priced in asset markets (Bolton and Kacperczyk, 2021; Pástor et al., 2022), so that how abstracts present the topic, as risks to be catalogued or as solutions to be implemented, is a meaningful descriptor of the field. Second, indicators built on citation databases depend on the database and on the way a corpus is assembled, which calls for checks of the data source before any interpretation. Existing bibliometric work mainly describes productivity, collaboration and thematic maps. Studies of the language of scientific abstracts show that tone changes over time (Vinkers et al., 2015). To our knowledge, few studies have jointly examined the bibliometric structure, the risk versus solution framing and the citation patterns of the literature on climate risk and financial stability. This statement rests on the works discussed in Section 2 and not on a systematic search of the literature, and it should be read with that caveat.

The objective of this study is to describe the bibliometric structure and the framing orientation of this literature and to examine how framing relates to citations. Four research questions guide the analysis. RQ1: what is the bibliometric structure of the corpus, and how consistent are its citation counts across Dimensions, Semantic Scholar and OpenAlex? RQ2: how well can a supervised classifier reproduce lexicon-based framing labels, and how do these labels compare with human coding on an audited sample? RQ3: how has the framing orientation of abstracts evolved between 2010 and 2025? RQ4: which thematic clusters can be identified, and do they differ in framing and in citation levels? Three propositions derived from the literature (Section 2.4) are examined descriptively: the share of solution-oriented abstracts increases over time (P1), framing is weakly related to citations (P2), and thematic clusters differ in framing (P3).

The study contributes in three ways. Empirically, it provides a descriptive map of a large corpus on climate risk and financial stability, with the citation, collaboration and thematic patterns that characterise it. Conceptually, it separates tone from framing orientation and treats the lexicon-based indicator as a thematic variable, which changes how its results should be read. Practically, it documents the limits of single-database citation counts and of lexicon-based silver labels for research evaluation. Checks of data quality, such as the comparison of citation sources, are reported as robustness tests and not as contributions.

The approach is computational and descriptive. It combines a transparent relevance screen, citation indicators, a lexicon-based framing indicator, supervised classifiers and lexical clustering, and it checks the framing indicator against a manual audit of 300 records. The study thus offers a replicable workflow for verifying that text-based indicators built from bibliographic exports measure what they claim to measure.

The remainder of the article is structured as follows. Section 2 reviews the literature and derives the propositions. Section 3 describes the data and the methods. Section 4 presents the results. Section 5 discusses the findings, their implications, the limitations and future research. Section 6 concludes.

2. Literature review

2.1 Climate risk and financial stability

Climate-related financial risk is usually organised in three categories. Physical risks come from acute weather events and chronic climatic shifts, transition risks from policy change, technological disruption and the repricing of carbon-intensive assets, and liability risks from litigation exposure (Task Force on Climate-related Financial Disclosures [TCFD], 2017; Christophers, 2017). Scenario work for central banks and supervisors translates these categories into macro-financial pathways (NGFS, 2021; Violle, 2025). The literature can be read through three transmission channels: asset markets, banks and investors, and the macro-prudential system.

In asset markets, the central question is whether carbon or transition risk is priced. Bolton and Kacperczyk (2021, 2023) find that firms with higher carbon emissions earn higher returns, which they read as a premium demanded for transition risk, and Reboredo and Ugolini (2022) relate transition risk to profitability and stock prices. Pástor et al. (2022) offer a competing reading in which green assets earn high realised returns because of unexpected increases in climate concerns, so that realised returns do not reveal expected returns. Ardia et al. (2023) and Engle et al. (2020) study how unexpected climate news and concerns move green and brown stocks and how this exposure can be hedged. Studies of attention and reaction also differ in what they stress: Choi et al. (2020) show that attention to climate change rises when local temperatures are abnormally high and that carbon-intensive firms underperform in such periods, Hong et al. (2019) find that markets underreact to drought risk in food company stocks, and Monasterolo and De Angelis (2020) examine the reaction of low-carbon and carbon-intensive equity indices to the Paris Agreement. Green bonds (Flammer, 2021) and the role of green investors as consumers (Sauzet and Zerbib, 2022) add a demand-side dimension, and Giglio et al. (2021) survey this asset-pricing literature.

Evidence on financial institutions is more fragmented. Institutional investors report that climate risks, and regulatory risks in particular, are material for their portfolio decisions (Krueger et al., 2020), text-based measures capture firm-level exposure to climate change (Sautner et al., 2023), and cap-and-trade regulation changes how banks allocate corporate credit (Ivanov et al., 2024). Measurement is itself a source of disagreement, since exposure can be proxied by emissions, by earnings-call text or by survey responses.

At the system level, financial network analysis shows that indirect exposures to fossil fuel assets can exceed direct holdings (Battiston et al., 2017), the European Systemic Risk Board formalised a "too late, too sudden" scenario in which delayed action forces an abrupt transition with systemic effects (Gros et al., 2016), and Dietz et al. (2016) estimate the expected climate value at risk of global financial assets and show that losses in the tail are much larger. Stress-testing frameworks (Acharya et al., 2023; Alogoskoufis et al., 2021), macro-financial models (Dafermos et al., 2018) and discussions of the role of central banks (Campiglio et al., 2018) develop these ideas, and the tail-risk logic builds on the measurement of systemic risk (Acharya et al., 2017). Adaptation choices under deep uncertainty are discussed in IPCC (2014) and Fankhauser et al. (1999).

2.2 Bibliometric studies in climate and sustainable finance

Citation analysis as a tool for describing scientific output goes back to Garfield (1972), and the use of several complementary indicators rather than a single one is a standard recommendation (Moed, 2005; Waltman, 2016). The h-index summarises productivity and impact in one number (Hirsch, 2005), but for a corpus assembled by a query it mainly reflects the size and construction of the corpus. Field normalisation is addressed by indicators such as the Field Citation Ratio provided by Dimensions (Hook et al., 2018), whose logic of comparing a paper with a

benchmark group is shared by the Relative Citation Ratio (Hutchins et al., 2016). Methodological guidance for bibliometric studies in management and business is provided by Zupic and Čater (2015) and Donthu et al. (2021), and software such as VOSviewer supports co-citation and co-occurrence mapping (Van Eck and Waltman, 2010), which this study does not perform. The choice of data source matters, since large comparisons show differences in coverage and metadata across Scopus, Web of Science, Dimensions, Crossref and Microsoft Academic (Visser et al., 2021), and OpenAlex has emerged as an open alternative (Priem et al., 2022). General treatments of the field are given by Milojević and Leydesdorff (2022). Bibliometric studies of sustainability research exist (for example Leal Filho et al., 2018), but this study did not conduct a systematic search for earlier bibliometric reviews of climate or sustainable finance, a limitation discussed in Section 5.3.

2.3 Framing, tone and text classification

Lexicon-based tone measurement has well-known domain limits. Loughran and McDonald (2011) show that almost three-quarters of the words that a general-purpose list classifies as negative are not negative in financial text, which motivated domain-specific vocabularies. In scientific writing, tone has been studied directly: Vinkers et al. (2015) document a marked increase in positive words in PubMed abstracts between 1974 and 2014. For climate finance abstracts, a further difficulty arises. Risk vocabulary ("risk", "crisis", "fragility") is the subject matter of many papers and does not necessarily express a pessimistic stance, while "green finance" or "net zero" signals a topic more than an opinion. The variable built in this study is therefore interpreted as a risk versus solution framing orientation, a thematic property of abstracts, and not as sentiment in the psychological sense.

Supervised classifiers trained on lexicon labels (silver or weak labels) inherit the biases of the lexicon, and a model that has the lexicon words among its features learns mainly to reproduce them. Linear support vector machines on TF-IDF features are a strong baseline for sparse text (Cortes and Vapnik, 1995; Joachims, 1998; Vapnik, 1999), and bigrams add local context, for instance distinguishing "transition opportunity" from "transition risk". They do not capture long-range syntax or semantics, and pre-trained language models such as FinBERT (Araci, 2019) and ClimateBERT (Webersinke et al., 2022) offer richer representations. Agreement between a classifier and the lexicon that produced its labels is evidence of internal consistency, not of validity, which requires comparison with human annotation.

2.4 Synthesis and propositions

The review shows three tensions. First, the literature on climate risk and financial stability is large and contested, in particular on whether transition risk is priced, so that a descriptive map of the field and of how it frames the topic is useful before specific hypotheses are tested. Second, bibliometric indicators depend on the database and on the way the corpus is built, which calls for checks of the data source before interpretation. Third, text-based indicators of framing depend on the definition of the construct and on labelling.

Three theoretical lenses guide the interpretation. The sociology of science predicts cumulative advantage in citations (Merton, 1968; Price, 1976), so that concentration of citations is expected in any field. Psychological research on negativity bias suggests that negative information receives more weight than positive information (Baumeister et al., 2001), which could favour attention to risk-oriented abstracts, whereas communication research on climate change suggests that solution-oriented messages engage audiences (Moser, 2014). These lenses lead to three propositions, examined descriptively and not tested causally. P1: the share of solution-oriented abstracts increases over time, in line with the policy attention to transition. P2: framing is related to citation counts, but the relation is weak, since cumulative advantage dominates.

P3: thematic clusters differ in framing, since topics that deal with risk measurement use risk vocabulary and topics that deal with green instruments use solution vocabulary.

3. Data and methodology

3.1 Corpus construction and relevance screening

Data were gathered from the Dimensions database (www.dimensions.ai), selected for its coverage, its provision of the Field Citation Ratio (FCR) and its export format. Query exports were merged (8,500 records) and deduplicated by Publication ID and DOI, which removed 500 duplicates and gave an initial export of 8,000 unique records published between 2010 and 2025 (Figure 1). The exports were limited by the export cap of the database and contain the same number of records for each publication year, so the export design and not the population of relevant publications fixes their annual size. Annual counts are therefore a product of the design and are not interpreted as publication trends. Records had to be Articles or Reviews according to Dimensions and to be written in English, and the queries combined climate risk keywords (climate risk, physical risk, transition risk, carbon risk) with financial stability keywords (financial stability, systemic risk, credit risk, financial markets). Dimensions also indexes other document types, so the corpus is described as articles and reviews and not as peer-reviewed publications. Missing journal names were recovered from the available metadata, and 120 records (2.9%) still have no journal name.

A transparent relevance screen was then applied to the title and abstract of each record. A record is retained if it contains at least one climate cue and at least one financial cue, and if the sum of the two cue counts is at least 2. Cue counts combine matches of 28 climate phrases and 38 financial phrases with overlaps between the text and two token lists (about 40 climate tokens and 50 financial tokens; the lists are available from the authors). The screen is deliberately permissive and keyword-based. It removed 3,836 of the 8,000 records (48.0%), which shows that the broad query keywords retrieved many records unrelated to climate and finance. A manual audit of 300 retained records (Sections 3.7 and 4.9) found that 57.3% are fully relevant and 73.7% fully or partly relevant, so that the retained corpus remains thematically broad. The analytical corpus contains 4,164 records, of which 87.6% have a relevance score of at least 4. All results below refer to this corpus unless stated otherwise. Table 1 summarises the design.

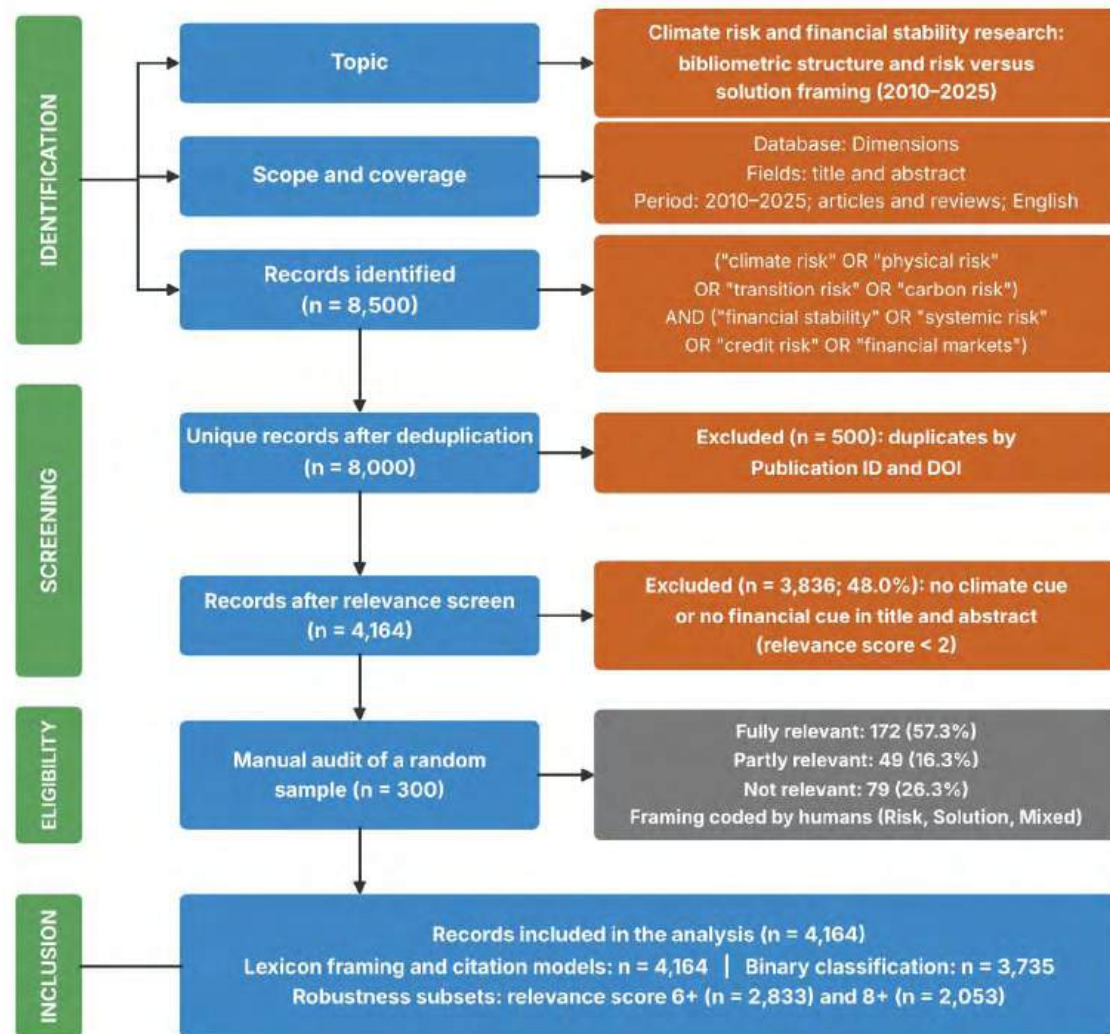
Table 1. Corpus design

Criterion	Description
Database	Dimensions
Initial export	8,000 unique records, 2010–2025, after removal of 500 duplicates (Figure 1)
Document type and language	Articles and reviews, English
Relevance screen	At least one climate cue and one financial cue (total of at least 2)
Records removed	3,836 (48.0%)
Analytical corpus	4,164 records
Binary classification sample	3,735 records labelled solution-oriented or risk-oriented
Clustering sample	4,164 records
Manual audit	300 random records (seed 42); first 100 with consensus labels

Note: Cue counts use title and abstract. The screen was audited on a sample of 300 retained records (Section 4.9).

Source: Authors' own elaboration.

Figure 1. Flow diagram of corpus construction, screening and audit



Note: Query terms as described in Section 3.1. Robustness subsets are defined by the relevance score of the screen (Section 4.10).

Source: Authors' elaboration from the Dimensions exports, the relevance screen and the manual audit.

3.2 Citation cross-validation across databases

To check the reliability of the Dimensions citation series, the DOI list of the initial export of 8,000 records (before the relevance screen) was submitted to two independent bibliographic APIs. The Semantic Scholar Academic Graph API returned citation counts for 7,393 records (92.5% of DOIs) and the OpenAlex API for 7,844 records (98.1%). The comparison used the Spearman rank correlation (Equation 1), where d_i is the difference between the ranks of record i in the two sources and n the number of matched records. These checks were not repeated on the 4,164 screened records, and the extraction dates of each source are not reported here. Rank correlations show that sources order papers consistently, but they do not show that absolute citation levels agree.

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)} \quad (1)$$

Table 2. Cross-database check of citation counts (initial export of 8,000 records)

Source	Matched records	Coverage of DOIs	Spearman ρ with Dimensions
Semantic Scholar	7,393	92.5%	0.933
OpenAlex	7,844	98.1%	0.976

Note: $p < 0.001$ for both coefficients. Computed before the relevance screen.

Source: Authors' calculations from Dimensions, Semantic Scholar and OpenAlex data.

3.3 Bibliometric performance analysis

Bibliometric performance was assessed with four indicators. The corpus h-index (Hirsch, 2005) is the largest integer h such that at least h papers have been cited at least h times (Equation 2), where $C(i)$ is the number of citations of the i -th most cited paper. Because the h-index of a query-based corpus depends on its size and construction, it describes the corpus and is not a measure of quality.

$$h = \max\{i \in \mathbb{N} : C(i) \geq i\} \quad (2)$$

The Field Citation Ratio, provided by Dimensions (Hook et al., 2018), compares the citations of a paper with those of papers of the same field, publication year and document type (Equation 3), where C_i is the number of citations of paper i and $C(\text{ref})$ the benchmark average computed by Dimensions.

$$\text{FCR}_i = \frac{C_i}{\bar{C}_{\text{ref}}(\text{field, year, type})} \quad (3)$$

Citation concentration is measured by the Gini coefficient derived from the Lorenz curve $L(x)$ of the citation distribution (Equation 4). The share of citations held by the top 1% of papers and the i10 index (number of papers with at least 10 citations) complete the indicators.

$$G = 1 - 2 \int_0^1 L(x) dx \quad (4)$$

3.4 Framing classification

3.4.1 Text preprocessing

Each document text is the concatenation of its title and abstract. The preprocessing lowercases the text, removes characters other than letters, spaces and hyphens (hyphens are kept to retain compound terms such as "net-zero" and "low-carbon"), removes 68 stop words and tokens shorter than three characters, and applies no stemming. All 4,164 documents contain text for classification and clustering.

3.4.2 Lexicon-based labelling

Framing labels are produced by a lexicon-based polarity score. The lexicon has a thematic part, with 20 solution-related words (for example green, transition, mitigation, adaptation, resilience, sustainable, renewable) and 10 risk-related words (for example risk, crisis, systemic, fragility, stranded, exposure, vulnerability), and a tonal part, with 37 positive and 52 negative evaluative words (for example benefit, growth, effective versus loss, damage, threat, decline). Because the thematic words name subject matter, the labels capture a risk versus solution framing orientation of the abstract and not the sentiment of the author. The polarity of document i is computed as (Equation 5), where P_i and N_i are the numbers of thematic and tonal positive and negative lexicon words in document i ; the score is set to 0 if no lexicon word is found. A tonal-only polarity, computed with the tonal words only, is kept as a comparison variable.

$$\text{POL}_i = \frac{P_i - N_i}{P_i + N_i} \quad (5)$$

Documents with $\text{POL} > 0.05$ are labelled solution-oriented, documents with $\text{POL} < -0.05$ risk-oriented, and the others mixed or neutral. The threshold of ± 0.05 was set a priori, and Section 4.10 reports results for ± 0.01 , ± 0.10 and ± 0.20 . This gives 2,365 solution-oriented, 1,370 risk-oriented and 429 mixed or neutral documents. The binary classification uses the 3,735 documents with a solution-oriented or risk-oriented label (63.3% solution-oriented).

3.4.3 Supervised classifiers

Three classifiers are compared: a linear SVM, a logistic regression (LR) and a random forest (RF). The 3,735 binary documents are split into 80% training ($n = 2,988$) and 20% hold-out test ($n = 747$) partitions, stratified by label, with a fixed random seed. The TF-IDF vectoriser is fitted on the training partition only and then applied to the test partition and to the full corpus. TF-IDF features use sublinear scaling (Equation 6):

$$\text{TF-IDF}(t, d) = [1 + \ln f(t, d)] \times \left[\ln \frac{1+|D|}{1+\text{df}(t)} + 1 \right] \quad (6)$$

where $f(t, d)$ is the frequency of term t in document d , $|D|$ the number of documents and $\text{df}(t)$ the number of documents containing t ; document vectors are then L2-normalised. The vectoriser keeps at most 5,000 features, unigrams and bigrams, with a minimum document frequency of 2 and a maximum document frequency of 85 per cent.

The SVM (scikit-learn LinearSVC) minimises a soft-margin squared hinge loss (Cortes and Vapnik, 1995) (Equation 7), where w is the weight vector, b the intercept, x_i the feature vector of document i , y_i its label (-1 or $+1$) and C the penalty parameter ($C = 1.0$):

$$\min_{w, b} \frac{1}{2} \|w\|^2 + C \sum_i [\max(0, 1 - y_i(w^T x_i + b))]^2 \quad (7)$$

Logistic regression models the probability of the solution-oriented class through a sigmoid link with L2 regularisation ($\lambda = 1.0$), with w and b the coefficients and the intercept (Equation 8):

$$P(y_i = 1 | x_i) = \frac{1}{1 + \exp(-(w^T x_i + b))} \quad (8)$$

The random forest aggregates $B = 100$ bootstrap decision trees by averaging their class probabilities, where $p_b(c | x)$ is the probability of class c given by tree b (Equation 9):

$$\hat{y}_{\text{RF}}(x) = \arg \max_{c \in \{0,1\}} \frac{1}{B} \sum_{b=1}^B p_b(c | x) \quad (9)$$

The SVM is also trained with class weights (class_weight = "balanced"). Hyperparameters were fixed a priori, with no search, and models were implemented with scikit-learn (Pedregosa et al., 2011). Cross-validated scores come from 5-fold stratified cross-validation on the training partition and test scores from the hold-out partition. Reported metrics are accuracy, F1 for the solution-oriented class, macro-averaged F1 and the Matthews correlation coefficient (MCC), which is independent of class size.

3.4.4 Agreement and ablation

Agreement between the lexicon labels and the SVM predictions is measured with Cohen's κ (Cohen, 1960) on the two classes (Equation 10), where p_o is the observed agreement and p_e the agreement expected by chance, and is read with the scale of Landis and Koch (1977):

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (10)$$

Since the SVM is trained on labels produced by the same lexicon, κ is computed both on the hold-out partition and, for the whole corpus, in-sample. Both values measure the internal consistency between two methods and not validity against human judgement. As an ablation, an SVM with unigram TF-IDF features is trained after removing all 119 lexicon words from the texts, to assess how much of the agreement depends on the lexicon words themselves.

3.5 K-Means topic clustering

The TF-IDF matrix of the 4,164 documents (unigrams and bigrams, 5,000 features, English stop words removed, minimum document frequency of 3) is reduced to 100 dimensions through truncated singular value decomposition, equivalent to latent semantic analysis (Equation 11), where U and Σ contain the first 100 singular vectors and values:

$$X_{\text{reduced}} = U_{100} \Sigma_{100} \quad (11)$$

K-Means is applied to the reduced matrix. The number of clusters is examined with a silhouette sweep from $k = 2$ to $k = 12$ (Rousseeuw, 1987), $k = 8$ is retained, and stability is assessed by

the adjusted Rand index (ARI) between the initial solution and solutions fitted on 30 bootstrap samples of 80% of the documents. Clusters larger than 18% of the corpus with fewer than 35% highly relevant documents would be split again, and a cluster audit reports size, relevance and framing. Cluster labels are interpretive descriptors based on dominant terms, and the clusters describe the lexical structure of abstracts and not an intellectual structure in the co-citation sense.

3.6 Multivariate model of citations

Citation counts are over-dispersed counts, so they are modelled with a negative binomial regression (Equation 12), where C_i is the number of citations of paper i , μ_i its expected value and α the dispersion parameter. Explanatory variables are the framing class (solution-oriented and mixed, with risk-oriented as reference), open access status (Green, Gold, Hybrid and Bronze, with closed access as reference), the logarithm of the number of authors and year fixed effects γ . Publication age is not entered, because with year fixed effects it is perfectly collinear. Standard errors are heteroskedasticity-robust (HC1), and a robustness model adds cluster fixed effects. Estimates are reported as incidence rate ratios (IRR). Section 4.10 reports further robustness checks: relevance thresholds of 6 and 8, polarity thresholds of ± 0.01 , ± 0.10 and ± 0.20 , fixed effects for the 53 journals with at least ten papers (other journals pooled, records without a journal name as a separate group) with standard errors clustered by journal, the Field Citation Ratio as an alternative outcome (OLS on $\ln(1 + \text{FCR})$), and interactions between framing and cluster, tested jointly with a Wald test. The K-Means groups are also compared with eight-topic NMF and LDA models fitted to the same texts, using the adjusted Rand index and normalised mutual information.

$$\ln \mu_i = \beta_0 + \beta_1 \text{Solution}_i + \beta_2 \text{Mixed}_i + \beta_3^T \mathbf{OA}_i + \beta_4 \ln \text{authors}_i + \gamma_{\text{year}(i)}, \quad \text{Var}(C_i) = \mu_i + \alpha \mu_i^2 \quad (12)$$

3.7 Manual audit of relevance and framing

A manual audit assessed the relevance of the retained records and the validity of the lexicon-based framing. A random sample of 300 records was drawn from the 4,164 (seed 42). Coders read the title and the abstract and did not see the lexicon or SVM labels. Relevance was coded as Yes (climate-related risk or policy and the financial system are both central in the abstract), Partial (one of the two is central and the other secondary) or No. Framing was coded as Risk, Solution or Mixed, according to whether the abstract mainly presents risks, losses and vulnerabilities, mainly presents solutions, instruments and mitigation or adaptation policies, or neither or both with equal weight.

One coder coded all 300 records. A second coder coded the first 100 records in a first round, which gave $\kappa = 0.22$ for relevance and $\kappa = 0.92$ for framing. The relevance rule was then made stricter and the disagreements were resolved by discussion, so that the first 100 records carry consensus labels and the remaining 200 carry the labels of the first coder. The final labels therefore do not provide an independent measure of inter-coder agreement. Precision is the share of audited records coded Yes (strict) or Yes or Partial (broad), reported with 95% Wilson confidence intervals. Agreement between human framing and the lexicon is measured with Cohen's κ on the three classes and on the records labelled Risk or Solution by both. As a test of the value of human labels, the TF-IDF SVM of Section 3.4 is retrained on the human framing labels (ternary, $n = 300$; binary Risk versus Solution, $n = 148$) in a nested cross-validation (5 folds repeated 3 times; an inner 3-fold search over C , n -gram range and maximum document frequency, selected on macro-F1), with the vectoriser fitted inside each training fold. On the binary task, the same SVM is also trained on sentence embeddings from two pre-trained language models, ClimateBERT (climatebert/distilroberta-base-climate-f; Webersinke et al., 2022) and FinBERT (ProsusAI/finbert; Araci, 2019), used without fine-tuning: each title and

abstract is truncated to 384 tokens, token vectors are mean-pooled, the embeddings are standardised inside each training fold, and the inner search covers C. The outer folds are the same as for the TF-IDF model. The majority class and the lexicon labels serve as baselines.

3.8 Sample sizes and reproducibility

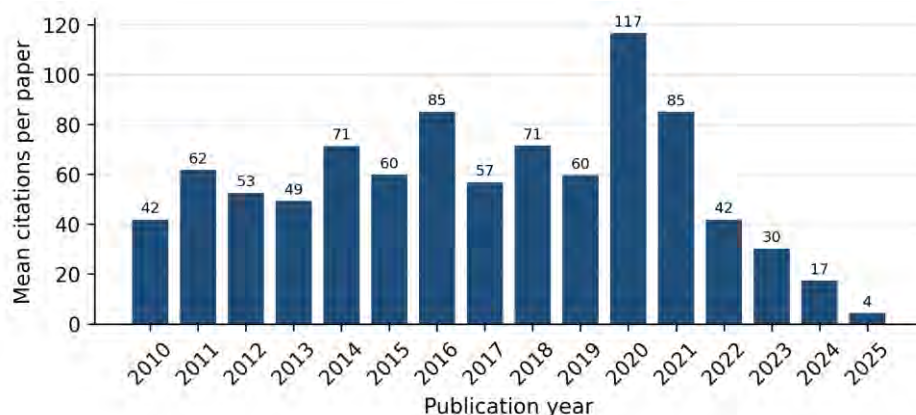
The sample size varies across analyses: 8,000 records in the initial export (and 7,393 and 7,844 matched in Semantic Scholar and OpenAlex), 4,164 in the analytical corpus, 3,735 with a binary framing label (2,988 for training and 747 for testing), 4,164 in the clustering and the regression, and 300 in the manual audit (148 with a Risk or Solution human label). Random seeds are fixed at 42, and the code and the results workbook are available from the authors on request.

4. Results

4.1 Bibliometric portrait of the corpus

Citation intensity peaks for the 2020 cohort (mean = 116.6 citations per paper) and falls for recent cohorts, which have had less time to accumulate citations (Figure 2). The 2021 cohort has a mean of 85.0, the 2016 cohort 85.1, and the 2024 and 2025 cohorts only 17.3 and 4.3. The corpus as a whole has a mean of 54.1 and a median of 17 citations per paper. The 2020 peak may reflect a small number of highly cited papers, but this was not tested. Because the initial export had the same number of records for each year, Figure 2 describes citation intensity by cohort and not publication output.

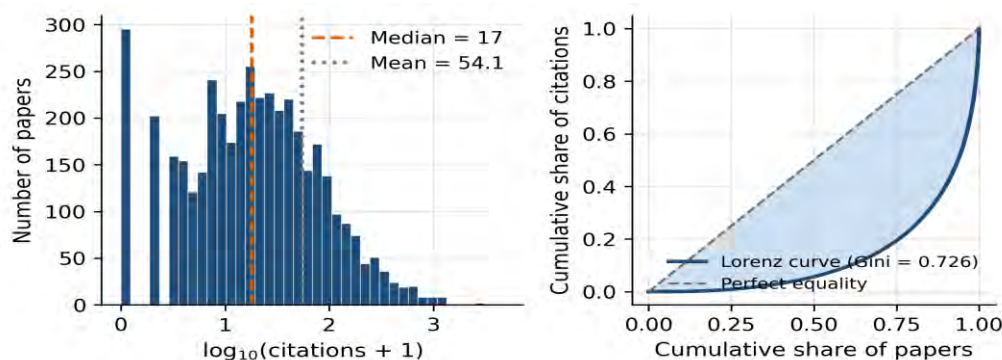
Figure 2. Mean citations per paper by publication year, 2010–2025



Note: Later cohorts have had less time to accumulate citations. The initial export had the same number of records for each year, so counts are not publication trends.

Source: Authors' calculations from Dimensions data (n = 4,164).

Figure 3. Citation distribution and Lorenz curve



Note: Gini coefficient = 0.726. The top 1% of papers (n = 42) hold 20.3% of citations. Corpus h-index = 210.

Source: Authors' calculations from Dimensions data.

The citation distribution is strongly right-skewed (Figure 3). The Gini coefficient is 0.726, and the top 1% of papers ($n = 42$, at least 596 citations) hold 20.3% of the 225,389 citations. This concentration is consistent with the cumulative advantage processes described by Merton (1968) and Price (1976). The corpus h-index is 210 and the i10 index is 2,645, so that 63.5% of the papers have at least 10 citations. These values describe the corpus and are not read as evidence of quality.

Table 3 presents citation statistics by open access status. Green open access papers have the highest mean (95.8) and median (39) citations, against a mean of 53.7 for closed access. Bronze (68.5) and Hybrid (64.6) lie in between, while Gold open access has the lowest mean (34.4) and median (11). No Diamond open access paper appears in the corpus. These descriptive comparisons control neither for publication year nor for journal or field, and the multivariate model of Section 4.7 revisits them.

Table 3. Citation statistics by open access status

Open access status	Papers (n)	Mean citations	Median citations
Green	254	95.8	39
Bronze	399	68.5	22
Hybrid	456	64.6	19
Closed access	2,031	53.7	18
Gold	1,024	34.4	11

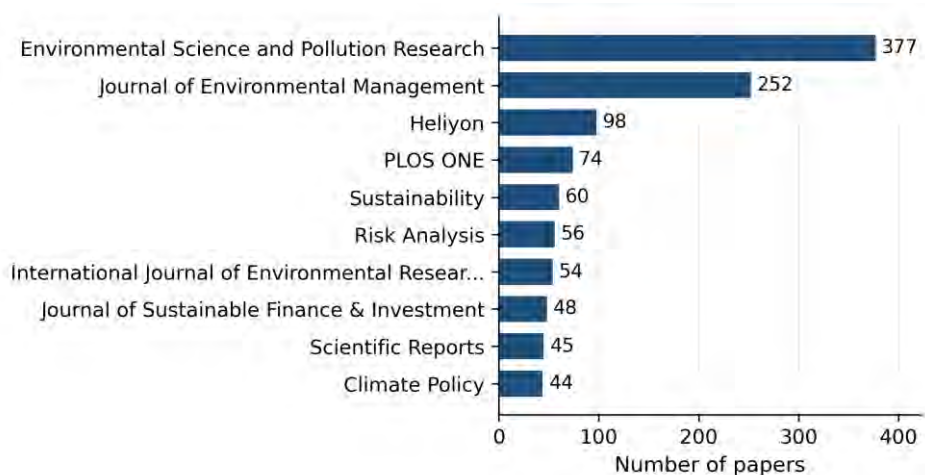
Note: Means and medians are not adjusted for publication year, journal or field.

Source: Authors' calculations from Dimensions data (n = 4,164).

4.2 Journals and collaboration

The corpus is concentrated in a small number of journals (Figure 4): Environmental Science and Pollution Research (377 papers, 9.1% of the corpus), Journal of Environmental Management (252, 6.1%) and Heliyon (98, 2.4%) lead, followed by PLOS ONE, Sustainability and Risk Analysis. Many leading outlets are broad environmental or multidisciplinary journals and not finance journals, which agrees with the thematic heterogeneity found in Section 4.8. Records without a journal name (120) are excluded from this count. Country information could not be derived reliably from the affiliation strings and is therefore not analysed.

Figure 4. Top 10 journals by number of papers

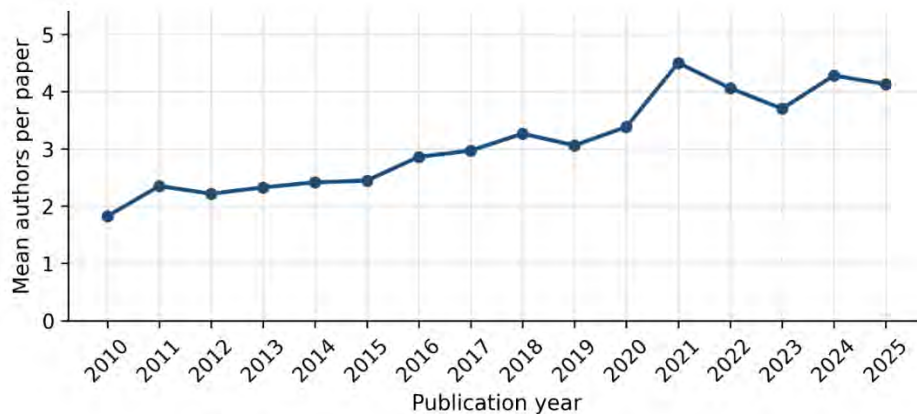


Note: Records without a journal name (n = 120) are excluded.

Source: Authors' calculations from Dimensions data.

Figure 5 shows a rise in the mean number of authors per paper from 1.8 in 2010 to 4.1 in 2025, with a peak of 4.5 in 2021. The rise is consistent with a general shift toward team-based science. The analysis uses means, which are sensitive to outliers (the maximum is 149 authors), and does not separate document types or test the factors behind the rise.

Figure 5. Mean number of authors per paper by publication year



Note: Means are sensitive to outliers; the maximum is 149 authors.

Source: Authors' calculations from Dimensions data.

4.3 Classifier performance

Table 4 reports classifier performance under 5-fold stratified cross-validation and on the hold-out test partition ($n = 747$). Cross-validated accuracy is 0.893 for the unweighted SVM, 0.893 for the class-weighted SVM, 0.869 for logistic regression and 0.861 for the random forest (Figure 6). On the test partition the class-weighted SVM achieves the highest accuracy (0.894), F1 (0.915) and MCC (0.776), and the unweighted SVM reaches 0.885, 0.909 and 0.752. The two SVM variants differ by less than the standard deviation of the cross-validation (about 0.018), so their ranking is not conclusive. Logistic regression (MCC = 0.717) and the random forest (MCC = 0.697) perform lower.

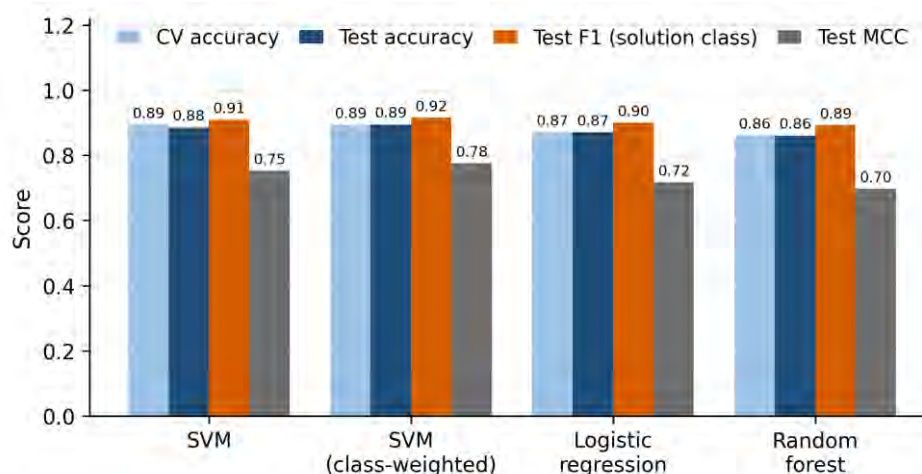
Table 4. Classifier performance (cross-validation and hold-out test)

Classifier	CV accuracy (SD)	CV F1	Test accuracy	Test F1	Test macro F1	Test MCC
SVM (unweighted)	0.893 (0.018)	0.917	0.885	0.909	0.876	0.752
SVM (class-weighted)	0.893 (0.017)	0.916	0.894	0.915	0.887	0.776
Logistic Regression	0.869 (0.021)	0.902	0.870	0.900	0.858	0.717
Random Forest	0.861 (0.020)	0.895	0.861	0.892	0.848	0.697

Note: CV = 5-fold stratified cross-validation on the training partition ($n = 2,988$); test = hold-out partition ($n = 747$). F1 refers to the solution-oriented class. Labels are lexicon-based, so scores measure the reproduction of the lexicon.

Source: Authors' calculations with scikit-learn.

Figure 6. Cross-validation and hold-out performance by classifier

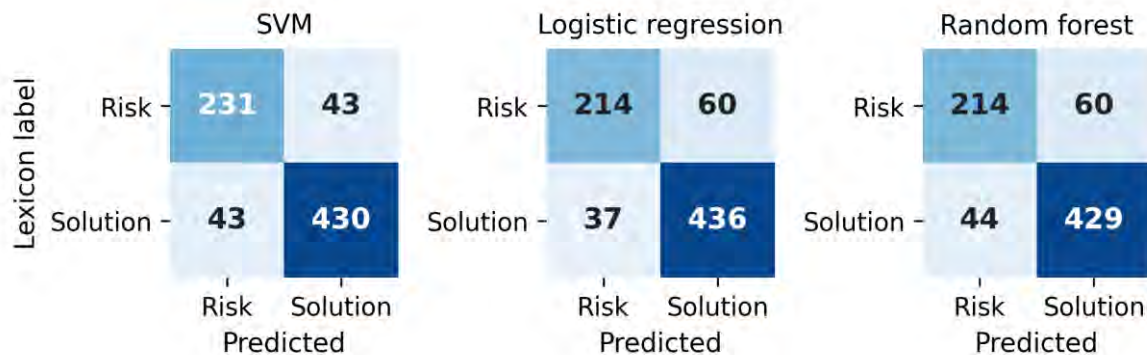


Note: F1 refers to the solution-oriented class. CV = cross-validation.

Source: Authors' calculations with scikit-learn.

For the unweighted SVM, precision and recall are both 0.909 for the solution-oriented class and 0.843 for the risk-oriented, minority class (Figure 7), so that the risk-oriented class is classified slightly less well. These scores measure how well each model reproduces lexicon-based labels. They are not accuracy against human annotation, since the features include the words that generated the labels.

Figure 7. Hold-out confusion matrices



Note: Test partition ($n = 747$). Rows are lexicon labels and columns are predicted labels.

Source: Authors' calculations with scikit-learn.

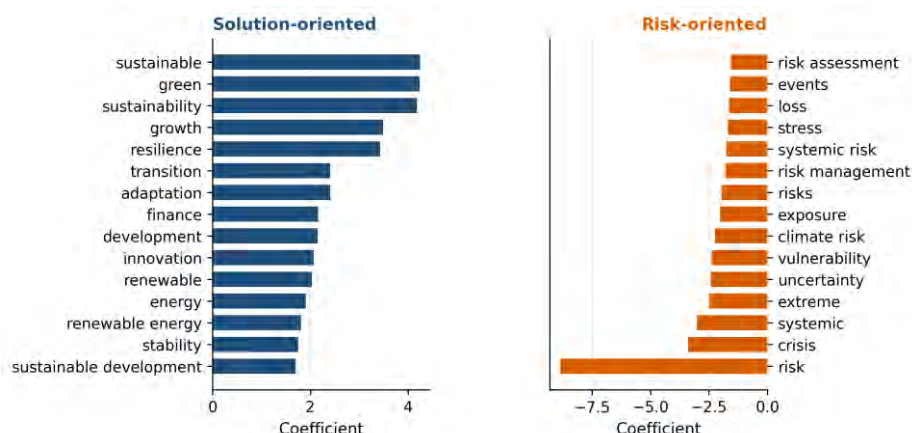
4.4 Agreement between lexicon labels and SVM predictions

Between the lexicon labels and the unweighted SVM predictions, Cohen's κ is 0.752 on the hold-out partition, which falls in the Substantial category of the Landis and Koch (1977) scale. Computed in-sample on the 3,735 binary documents, the observed agreement is 97.7% (1,327 risk-oriented and 2,321 solution-oriented documents in agreement, 87 disagreements) and κ is 0.950. The higher in-sample value reflects the use of the same documents for training.

The ablation shows how much of this agreement depends on the lexicon words. When all 119 lexicon words are removed from the texts, a unigram SVM reaches an accuracy of 0.739, an F1 of 0.797, a macro F1 of 0.716 and an MCC of 0.432. This is well above chance, which shows that the vocabulary around the lexicon words carries information on framing, but it is far below the 0.885 obtained with the lexicon words present (the ablated model also uses unigrams only). The agreement between the lexicon and the SVM therefore reflects mainly the lexicon words themselves, and it indicates internal consistency and not validity.

4.5 Most discriminative features

Figure 8. Top 15 most discriminative TF-IDF features by class



Note: Features ranked by logistic regression coefficients. Most terms belong to the lexicon.

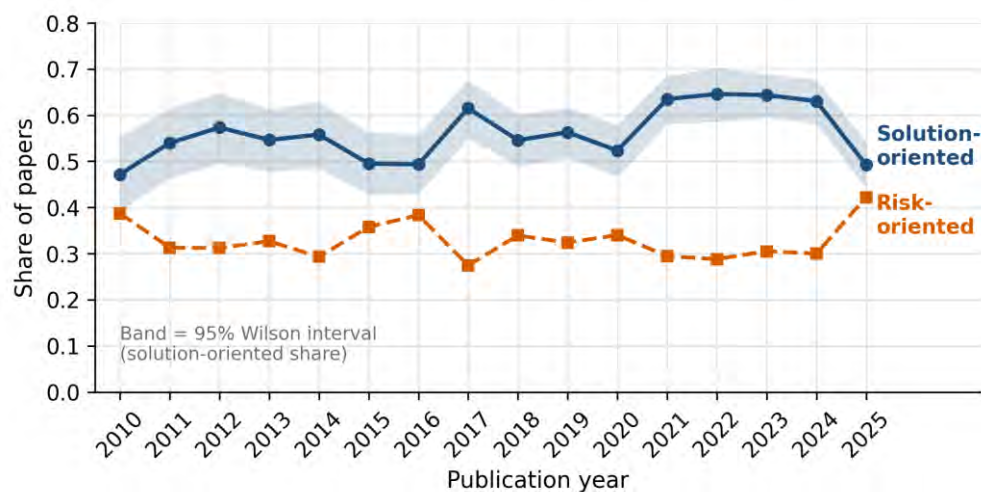
Source: Authors' calculations.

Figure 8 shows the 15 most discriminative TF-IDF features for each class, measured by logistic regression coefficients. The solution-oriented class is characterised by terms such as "sustainable", "green", "sustainability", "growth" and "resilience", and the risk-oriented class by "risk" (the largest coefficient in absolute value, -8.9), "crisis", "systemic", "extreme", "uncertainty" and "vulnerability". Most of these terms belong to the lexicon, so the classifier learns largely the lexicon, as the ablation indicates. They are topical, which supports reading the indicator as a thematic framing variable: abstracts on risk measurement necessarily use risk vocabulary (Acharya et al., 2017; Brunnermeier and Pedersen, 2009).

4.6 Framing over time

Figure 9 shows the annual share of solution-oriented and risk-oriented abstracts, with 95% Wilson confidence intervals. Across the corpus, 56.8% of abstracts are solution-oriented and 32.9% are risk-oriented. The annual share of solution-oriented abstracts ranges from 47.2% in 2010 to 64.6% in 2022, and it is between 63.0% and 64.6% from 2021 to 2024, then drops to 49.2% in 2025, when the risk-oriented share reaches 42.2%.

Figure 9. Annual share of solution-oriented and risk-oriented abstracts



Note: The remainder to 100% is mixed or neutral abstracts. Annual sample sizes range from 142 to 390.

Source: Authors' calculations ($n = 4,164$).

A Mann-Kendall test on the 16 annual shares gives $\tau = 0.30$ ($S = 36$, $p = 0.115$), so that no monotonic trend is established. Excluding the incomplete 2025 cohort, $\tau = 0.47$ ($p = 0.018$), which depends on this choice. A logistic regression of the framing label on publication year at document level gives an odds ratio of 1.008 per year ($p = 0.28$). P_1 is therefore not supported: the data show a higher share of solution-oriented abstracts from 2021 to 2024 than before 2021, and a lower share in 2025, but no structural reorientation. The 2025 value is not explained here, and the wide intervals (± 5 points) show that year-to-year differences are of the order of sampling error for several years.

4.7 Framing, citations and collaboration

Table 5 reports Spearman rank correlations. Citations and the Field Citation Ratio are strongly correlated ($\rho = 0.794$), as expected, since FCR is a normalised citation count. Thematic polarity is only weakly associated with citations ($\rho = 0.055$, $p < 0.001$) and FCR (0.068), and tonal polarity is not associated with citations ($\rho = 0.014$, $p = 0.36$). The number of authors ($\rho = 0.190$) and the age of the paper (0.208) are more strongly associated with citations, and authors are negatively correlated with age (-0.322), which reflects the rise of team science. With 4,164

observations, very small coefficients can be statistically significant, so that significance is not interpreted as importance.

Table 5. Spearman rank correlations between citations, framing and collaboration

	Citations	FCR	Thematic polarity	Tonal polarity	N authors	Age
Citations	1.000	0.794	0.055	0.014	0.190	0.208
FCR	0.794	1.000	0.068	0.036	0.142	0.070
Thematic polarity	0.055	0.068	1.000	0.625	-0.031	-0.016
Tonal polarity	0.014	0.036	0.625	1.000	-0.029	0.002
N authors	0.190	0.142	-0.031	-0.029	1.000	-0.322
Age (years)	0.208	0.070	-0.016	0.002	-0.322	1.000

Note: Thematic polarity is the polarity score of Equation 5 and tonal polarity uses the tonal words only.

Source: Authors' calculations ($n = 4,164$).

Table 6 reports the negative binomial model. After controlling for year, open access status and the number of authors, solution-oriented abstracts have 26.5% higher expected citations than risk-oriented abstracts (IRR = 1.265; 95% CI 1.094 to 1.463; $p = 0.002$), and mixed or neutral abstracts do not differ from risk-oriented ones (IRR = 0.960). Doubling the number of authors is associated with about 39% more citations (IRR of the logarithm = 1.613). Green open access is associated with 31% more citations than closed access (IRR = 1.310; $p = 0.014$) and Gold open access with 27% fewer (IRR = 0.732; $p = 0.001$), while Hybrid and Bronze do not differ from closed access. When cluster fixed effects are added, the association of solution-oriented framing is smaller but remains (IRR = 1.179; 95% CI 1.014 to 1.372; $p = 0.033$), and Green open access (IRR = 1.464) and Gold (IRR = 0.796) keep their signs.

Table 6. Negative binomial regression of citation counts

Variable	Model 1: IRR [95% CI]	p	Model 2: IRR [95% CI]	p
Solution-oriented abstract	1.265 [1.094, 1.463]	0.002	1.179 [1.014, 1.372]	0.033
Mixed or neutral abstract	0.960 [0.786, 1.172]	0.688	0.948 [0.784, 1.146]	0.582
Green open access	1.310 [1.057, 1.624]	0.014	1.464 [1.177, 1.819]	< 0.001
Gold open access	0.732 [0.606, 0.885]	0.001	0.796 [0.656, 0.966]	0.021
Hybrid open access	1.144 [0.938, 1.394]	0.183	1.304 [1.055, 1.612]	0.014
Bronze open access	1.019 [0.835, 1.242]	0.856	1.141 [0.932, 1.397]	0.202
ln(number of authors)	1.613 [1.480, 1.757]	< 0.001	1.595 [1.459, 1.745]	< 0.001
Dispersion α (model 1)	1.667 (SE 0.051)			

Note: Incidence rate ratios (IRR) with 95% confidence intervals from heteroskedasticity-robust standard errors (HC1). Reference categories: risk-oriented abstracts and closed access. Year fixed effects are included but not shown. Model 2 adds cluster fixed effects.

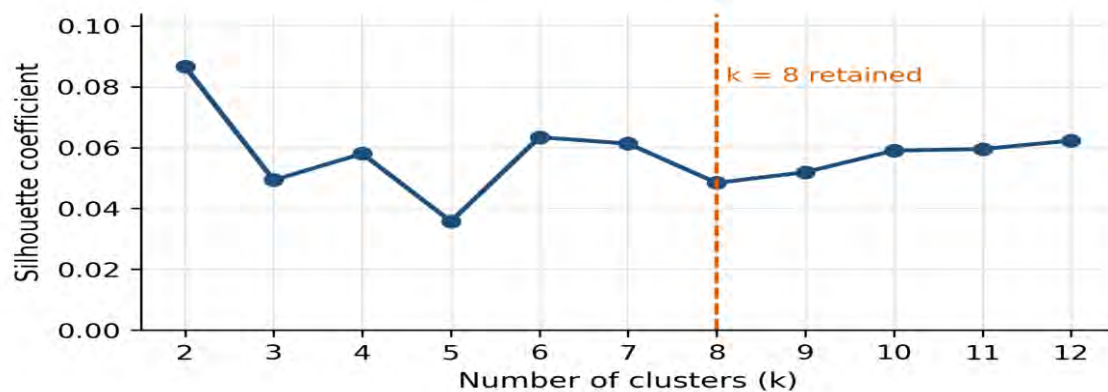
Source: Authors' calculations ($n = 4,164$).

P2 is therefore only partly supported. The bivariate rank association is weak, but conditional on year, open access and team size solution-oriented abstracts have higher citation counts. These are associations: the framing variable is lexicon-based, the association weakens when journal and cluster fixed effects are both included (Section 4.10), and the results do not show that framing causes citations. The positive association is not consistent with a citation advantage of risk-oriented abstracts (Baumeister et al., 2001), and it is compatible with the idea that solution-oriented topics such as green finance attract citations (Moser, 2014), but this was not tested.

4.8 Lexical clustering of abstracts

Figure 10 shows the silhouette coefficient for k from 2 to 12. The values are low throughout (between 0.036 and 0.087), with the highest value at $k = 2$ (0.087), 0.063 at $k = 6$ and 0.048 at $k = 8$, which indicates a weak cluster structure. The bootstrap ARI of the retained eight-cluster solution is 0.55 (SD = 0.11), which indicates moderate stability. The value $k = 8$ was retained as a descriptive choice that separates thematic groups and avoids one very large cluster, not as a sharply identified number of topics. The splitting rule was not triggered, as no large cluster had a low share of highly relevant documents.

Figure 10. Silhouette coefficient by number of clusters

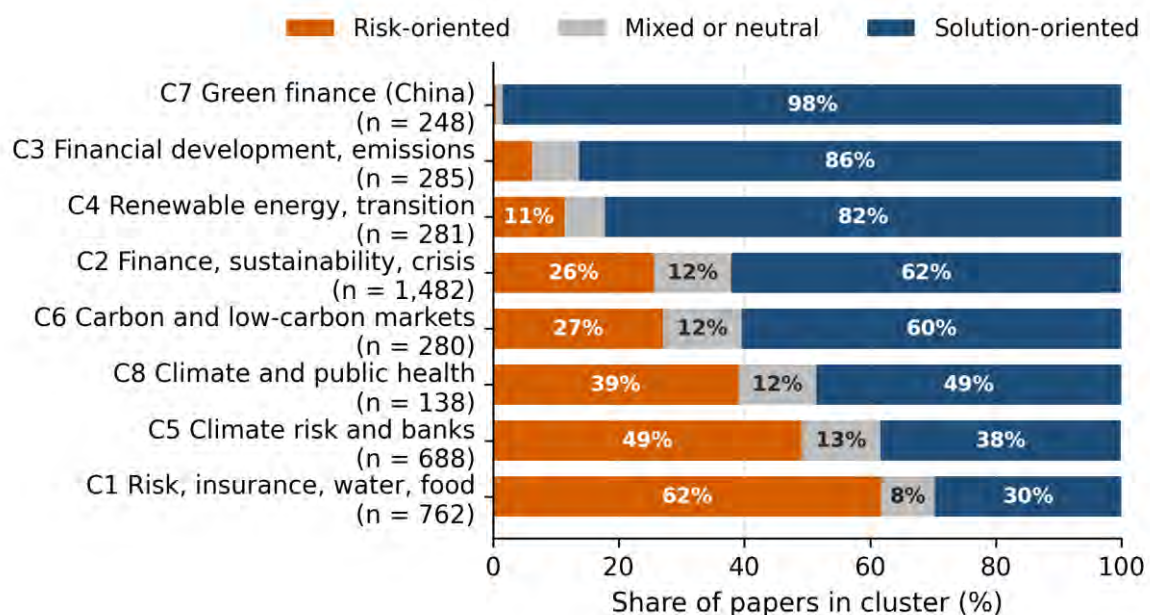


Note: Silhouette computed on the 100-dimensional SVD embedding; $k = 8$ retained.

Source: Authors' calculations.

Table 7 and Figure 11 show a bipolar distribution of framing across clusters, which supports P3. Green finance (C7, 248 papers) has 98.4% of its abstracts classified as solution-oriented, and financial development and emissions (C3, 285 papers) 86.3%, followed by renewable energy and transition (C4, 82.2%). Risk, insurance, water and food (C1, 762 papers) has 61.8% of risk-oriented abstracts, and climate risk and banks (C5, 688 papers) 49.1%. Because these clusters are formed from lexical similarity and the framing indicator comes from a lexicon that contains words such as "risk" and "green", part of this association is mechanical.

Figure 11. Framing composition by thematic cluster



Note: Clusters ordered by share of solution-oriented abstracts. Labels are interpretive.

Source: Authors' calculations.

The clusters also show how broad the corpus is. The core topic, climate risk and banks (C5), represents 16.5% of the corpus, whereas C1 and C2, flagged as residual or heterogeneous in the audit, hold 18.3% and 35.6%. Mean citations are highest for C3 (105.4) and lowest for C5 (46.5; median 12), but cohort composition may explain part of these differences, since mean age varies from 4.2 years (C7) to 8.8 years (C2). Differences in citations between clusters were not tested for significance.

Table 7. Thematic cluster profiles by framing and citations

Cluster	Papers	% of corpus	Mean citations	Median citations	Solution (%)	Risk (%)	Status
C1 Risk, insurance, water, food	762	18.3	54.8	20.0	29.7	61.8	Heterogeneous
C2 Finance, sustainability, crisis	1,482	35.6	47.8	14.0	62.0	25.6	Heterogeneous
C3 Financial development, emissions	285	6.8	105.4	41.0	86.3	6.3	Core
C4 Renewable energy, transition	281	6.7	47.8	23.0	82.2	11.4	Core
C5 Climate risk and banks	688	16.5	46.5	12.0	38.2	49.1	Core
C6 Carbon and low-carbon markets	280	6.7	52.0	20.5	60.4	27.1	Core
C7 Green finance (China)	248	6.0	54.1	17.0	98.4	0.4	Core
C8 Climate and public health	138	3.3	68.1	15.5	48.6	39.1	Core

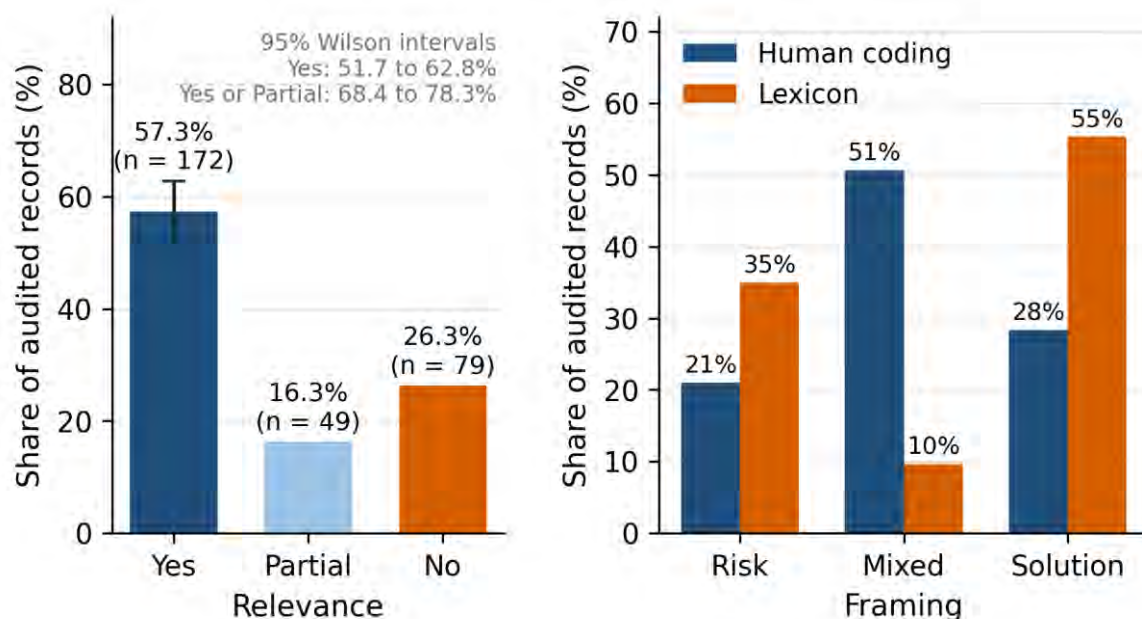
Note: Cluster labels are interpretive, based on dominant terms. Status is the result of the cluster audit (residual or heterogeneous versus core). Mixed or neutral abstracts are not shown (the remainder to 100%).

Source: Authors' calculations.

4.9 Manual audit of relevance and framing

Table 8 and Figure 12 summarise the audit. Of the 300 audited records, 172 (57.3%; 95% CI 51.7 to 62.8) are fully relevant and 221 (73.7%; 95% CI 68.4 to 78.3) are fully or partly relevant, whereas 79 (26.3%) are not relevant. About a quarter of the retained corpus is therefore off-topic even after the screen. The relevance score of the screen separates relevant from non-relevant records only moderately (AUC = 0.76). Raising the threshold from 2 to 6 would keep 2,833 records with 67.9% fully relevant, losing 16% of the fully relevant audited records, and a threshold of 8 would keep 2,053 records with 76.0% fully relevant, losing 32%. These trade-offs are estimated on 300 records, and the threshold of 2 was retained for the analyses of Sections 4.1 to 4.8, which therefore describe a broad corpus that includes off-topic records; Section 4.10 repeats the main analyses on the stricter subsets.

Figure 12. Relevance and framing in the manual audit (n = 300)



Note: Left: relevance coded by the human coder. Right: framing coded by the human coder and assigned by the lexicon. The first 100 records carry consensus labels.

Source: Authors' calculations from the manual audit.

Table 8. Manual audit of 300 retained records: relevance and framing

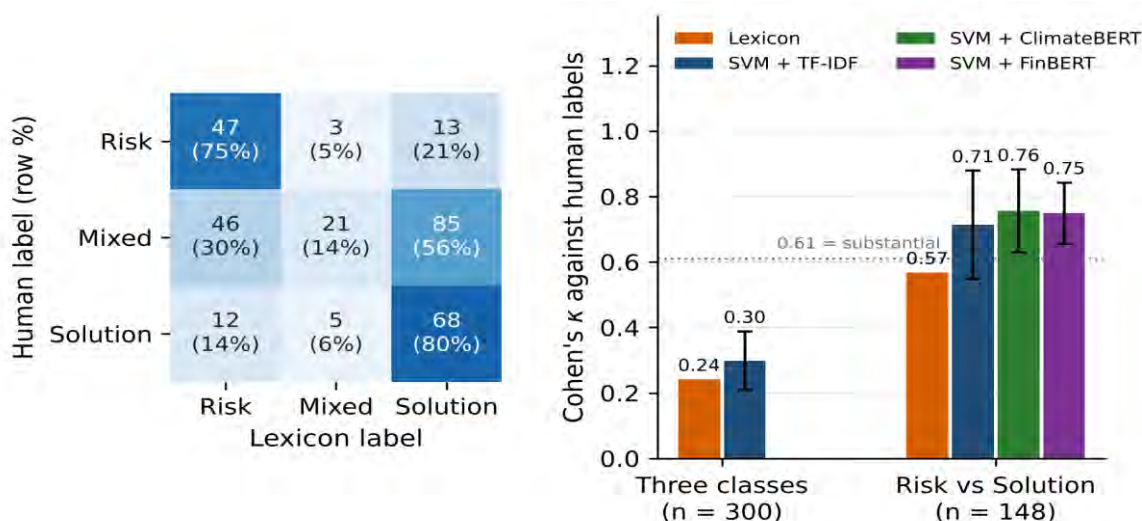
Measure	Result
Fully relevant (Yes)	172 of 300 = 57.3% [51.7, 62.8]
Fully or partly relevant (Yes or Partial)	221 of 300 = 73.7% [68.4, 78.3]
Not relevant (No)	79 of 300 = 26.3%
Human framing, all audited records (n = 300)	Mixed 50.7%, Solution 28.3%, Risk 21.0%
Lexicon framing, all audited records (n = 300)	Mixed 9.7%, Solution 55.3%, Risk 35.0%
κ human versus lexicon, three classes	0.242 (agreement 45.3%)
κ human versus lexicon, Risk versus Solution (n = 140)	0.635 (agreement 82.1%)
κ human versus SVM trained on lexicon, Risk versus Solution (n = 140)	0.651

Note: Wilson 95% confidence intervals in brackets. The first 100 records carry consensus labels (Section 3.7). Human framing among the 221 fully or partly relevant records: Mixed 50.7%, Solution 29.9%, Risk 19.5%.

Source: Authors' calculations from the manual audit.

Human framing differs markedly from lexicon framing (Figures 12 and 13). Among the 300 audited records, the human coder labels 50.7% Mixed, 28.3% Solution and 21.0% Risk (50.7%, 29.9% and 19.5% among the 221 fully or partly relevant records), whereas the lexicon labels only 9.7% Mixed, 55.3% Solution and 35.0% Risk. Of the 152 records coded Mixed by the human coder, the lexicon assigns 85 to Solution and 46 to Risk, and only 21 to Mixed. Cohen's κ between human and lexicon framing is 0.242 on three classes (agreement 45.3%) and 0.635 on the 140 records that both label Risk or Solution (agreement 82.1%). The SVM trained on the lexicon labels agrees with the human labels at a similar level on the binary task ($\kappa = 0.651$; $n = 140$). The lexicon therefore separates risk-oriented from solution-oriented abstracts reasonably well, but it over-assigns a direction to abstracts that human readers judge to have no clear orientation.

Figure 13. Human framing compared with the lexicon and with SVMs trained on human labels



Note: Left: human label (rows) by lexicon label (columns), counts and row percentages. Right: Cohen's κ against human labels; SVM trained on human labels with TF-IDF, ClimateBERT or FinBERT representations, in nested cross-validation (mean and standard deviation over 15 outer folds). The lexicon κ in the binary panel counts a lexicon label of Mixed as a disagreement.

Source: Authors' calculations from the manual audit.

Table 9 and Figure 13 report the SVM retrained on human labels. On the binary task the SVM reaches $\kappa = 0.713$ (SD = 0.171; macro-F1 = 0.856; $n = 148$), against 0.568 for the lexicon labels on the same records (where a lexicon label of Mixed counts as an error). On the ternary task the SVM reaches $\kappa = 0.298$ (macro-F1 = 0.542), against 0.242 for the lexicon, and its accuracy of 0.588 only moderately exceeds the 0.507 of the majority class. With pre-trained embeddings, the SVM reaches $\kappa = 0.756$ (SD = 0.127; accuracy = 0.878; macro-F1 = 0.876) with

ClimateBERT and $\kappa = 0.749$ (SD = 0.094; accuracy = 0.876; macro-F1 = 0.874) with FinBERT, against $\kappa = 0.713$ (accuracy = 0.861) with TF-IDF. The gain of about 0.04 in κ is smaller than the standard deviation across folds, so that the three representations perform similarly on this sample, although the embeddings give less variable results. The improvement over the lexicon is clear in sign but its uncertainty is large (the SD of κ is 0.09 to 0.18 across the outer folds), the Risk class has only 63 records, and the sample was labelled mainly by one coder, so that the results are indicative.

Table 9. SVM retrained on human framing labels with TF-IDF and pre-trained embeddings (nested cross-validation)

Task and representation	n	Majority class κ	Lexicon κ	Model κ (SD)	Model macro-F1
Ternary, TF-IDF	300	0.000	0.242	0.298 (0.093)	0.542
Binary, TF-IDF	148	0.000	0.568	0.713 (0.171)	0.856
Binary, ClimateBERT embeddings	148	0.000	0.568	0.756 (0.127)	0.876
Binary, FinBERT embeddings	148	0.000	0.568	0.749 (0.094)	0.874
Binary, relevant records, TF-IDF	109	0.000	0.502	0.731 (0.178)	0.864

Note: κ is Cohen's kappa against the human labels (mean and standard deviation over 15 outer folds). The model is a linear SVM; ClimateBERT and FinBERT are used as frozen mean-pooled embeddings with identical outer folds. The binary rows use the records labelled Risk or Solution by the human coder ($n = 148$); Table 8 reports κ on the 140 of them that the lexicon also labels Risk or Solution. The lexicon baseline is scored against the human labels on the same records, a lexicon label of Mixed being counted as a disagreement in the binary rows.

Source: Authors' calculations with scikit-learn.

4.10 Robustness and sensitivity

Table 10 reports how the main descriptive and citation results change with the relevance threshold of the screen and with the polarity threshold of the lexicon. Raising the relevance threshold to 6 ($n = 2,833$) or 8 ($n = 2,053$), which raises the share of fully relevant records from 57.3% to about 68% and 76% according to the audit, slightly increases the share of solution-oriented abstracts (from 56.8% to 60.1% and 60.8%) and leaves the citation association unchanged (IRR = 1.267 and 1.283). Changing the polarity threshold from ± 0.01 to ± 0.20 moves records from the two directional classes into the mixed class (mixed share from 10.2% to 19.3%), but the IRR of solution-oriented framing stays between 1.26 and 1.30 and is significant at the 1% level in all variants. In no variant does the Mann-Kendall test establish a monotonic trend in the annual share of solution-oriented abstracts (p between 0.08 and 0.44), so that the conclusion on P1 does not depend on these choices.

Table 10. Sensitivity of framing shares, time trend and citation association to the relevance and polarity thresholds

Variant	n	Solution / Risk / Mixed (%)	Mann-Kendall τ (p)	IRR solution-oriented [95% CI]	p
Relevance score 2+ (baseline)	4,164	56.8 / 32.9 / 10.3	0.30 (0.115)	1.265 [1.094, 1.463]	0.002
Relevance score 6+	2,833	60.1 / 31.8 / 8.1	0.22 (0.260)	1.267 [1.099, 1.461]	0.001
Relevance score 8+	2,053	60.8 / 31.7 / 7.5	0.27 (0.163)	1.283 [1.089, 1.510]	0.003
Polarity threshold ± 0.01	4,164	56.9 / 33.0 / 10.2	0.33 (0.079)	1.264 [1.094, 1.461]	0.002
Polarity threshold ± 0.05 (baseline)	4,164	56.8 / 32.9 / 10.3	0.30 (0.115)	1.265 [1.094, 1.463]	0.002
Polarity threshold ± 0.10	4,164	55.7 / 32.0 / 12.3	0.27 (0.163)	1.299 [1.120, 1.506]	< 0.001
Polarity threshold ± 0.20	4,164	52.1 / 28.6 / 19.3	0.15 (0.444)	1.291 [1.102, 1.511]	0.002

Note: Relevance rows use the polarity threshold ± 0.05 and polarity rows the relevance threshold 2. τ and p from the Mann-Kendall test on the 16 annual shares of solution-oriented abstracts. IRR from the negative binomial model of Table 6 (year fixed effects, open access, ln authors; HC1 standard errors).

Source: Authors' calculations.

Table 11 reports alternative specifications of the citation model. Adding fixed effects for the 53 journals with at least ten papers and clustering standard errors by journal reduces the IRR to 1.206 (95% CI 1.022 to 1.424; $p = 0.027$), and adding cluster fixed effects as well reduces it to

1.169 (0.999 to 1.369; $p = 0.051$). With the field-normalised FCR as outcome, solution-oriented abstracts have a 10.2% higher FCR ($p = 0.014$). Restricting the sample to records with a relevance score of 8 or more, with journal fixed effects, gives an IRR of 1.308 ($p = 0.005$). The association therefore holds under the relevance and framing thresholds and with the alternative outcome, but it weakens when journal and cluster are both controlled, which indicates that part of it reflects where and on which topics solution-oriented papers are published. A Wald test of the interactions between framing and cluster rejects equal associations across clusters ($\chi^2 = 22.95$, 7 df, $p = 0.002$): the IRR of solution-oriented framing is above 1 in six clusters (1.05 to 1.45) and below 1 in climate risk and banks (C5, 0.69) and green finance (C7, 0.71).

Table 11. Alternative specifications of the citation model

Specification	Outcome	n	Estimate (solution-oriented)	95% CI	p
Baseline: year FE, open access, ln authors; HC1	Citations (IRR)	4,164	1.265	[1.094, 1.463]	0.002
+ journal FE; SE clustered by journal	Citations (IRR)	4,164	1.206	[1.022, 1.424]	0.027
+ journal FE and cluster FE; SE clustered by journal	Citations (IRR)	4,164	1.169	[0.999, 1.369]	0.051
Baseline controls; HC1	ln(1 + FCR) (ratio)	4,164	1.102	[1.020, 1.191]	0.014
Relevance score 8+, with journal FE; SE clustered by journal	Citations (IRR)	2,053	1.308	[1.085, 1.577]	0.005

Note: Negative binomial models except the FCR row (OLS on $\ln(1 + \text{FCR})$, estimate reported as $\exp(b)$). Journal FE: the 53 journals with at least ten papers, other journals pooled and records without a journal name as a separate group. Reference category: risk-oriented abstracts.

Source: Authors' calculations.

Finally, the K-Means groups were compared with eight-topic NMF and LDA models fitted to the same texts, each abstract being assigned to its dominant topic. Agreement is moderate with NMF (ARI = 0.36; normalised mutual information 0.53) and low with LDA (ARI = 0.12; 0.23), and NMF and LDA agree little with each other (ARI = 0.14), which confirms that the thematic partition depends on the method. The contrast in framing between groups is, however, present under all three methods: the share of solution-oriented abstracts ranges from 29.7% to 98.4% across K-Means groups, from 12.1% to 94.7% across NMF topics and from 23.8% to 85.1% across LDA topics. The groups are therefore read as lexical groupings and not as a stable intellectual structure.

5. Discussion

5.1 Main findings

Regarding RQ1, the corpus is strongly concentrated in citations (Gini = 0.726), dominated by a few broad environmental journals, and shows consistent rankings across Dimensions, Semantic Scholar and OpenAlex on the initial export ($p = 0.933$ and 0.976). This consistency supports using Dimensions for relative comparisons, but it does not show that absolute levels agree (Visser et al., 2021). The relevance screen removed 48% of the initial export, which is a first result in itself: the keywords used for retrieval produced a large share of records unrelated to the topic, and the screen is keyword-based and was audited only on a sample of retained records.

Regarding RQ2, a linear SVM reproduces the lexicon labels with an accuracy of 0.885 (MCC = 0.752) on the hold-out partition, and κ is 0.752 on this partition. Since the SVM learns from the lexicon words, this agreement is expected, and the ablation shows that accuracy falls to 0.739 without them. The manual audit gives the relevant validity evidence: the lexicon agrees with human framing at $\kappa = 0.24$ on three classes and 0.64 on risk versus solution, because it

assigns a direction to many abstracts that human readers code as mixed. The indicator is therefore better read as a measure of the vocabulary of risk and solutions than as a measure of the stance of an abstract. A classifier trained on human labels does better on the binary task ($\kappa = 0.71$ with TF-IDF, 0.76 with ClimateBERT and 0.75 with FinBERT embeddings), but the differences between representations are within sampling variation and the human-coded set is small.

Regarding RQ3, the share of solution-oriented abstracts is above one half in the corpus (56.8%), and the annual share shows no stable monotonic trend, with a higher level from 2021 to 2024 and a lower one in 2025. P1 is not supported under any of the relevance and framing thresholds tested in Section 4.10. Regarding RQ4, eight clusters with very different framing are found, and the contrast also appears with NMF and LDA topics, which supports P3, but the structure is weak, the association between framing and cluster vocabulary is partly mechanical, and the over-assignment of direction by the lexicon may exaggerate the contrasts between clusters. Framing is weakly related to citations in rank terms, while solution-oriented abstracts have higher expected citations conditional on controls, so that P2 is partly supported. This association survives stricter relevance screens, alternative framing thresholds, journal fixed effects and a field-normalised outcome, but it weakens when journal and cluster effects are combined and it differs across clusters (Section 4.10). This pattern fits the dominance of cumulative advantage and team size in citation outcomes (Merton, 1968; Price, 1976), and it does not support a negativity advantage (Baumeister et al., 2001).

5.2 Implications

For research evaluation, the results illustrate why several indicators should be used together and why indicators should be interpreted in light of how a corpus is built (Hicks et al., 2015; Waltman, 2016; DORA, 2013). The h-index and mean citations of a query-based export describe the export, and field-normalised indicators are preferable for comparisons across cohorts. The open access results are mixed: Green open access is associated with more citations and Gold with fewer, even after controls, so that open access status should not be used as a proxy for quality or influence without further evidence.

For the use of text classification in bibliometrics, the main implication is that lexicon-derived labels and a classifier trained on them produce an internally consistent but unvalidated indicator, and that an ablation of the lexicon words is a useful check. Studies that rely on such indicators should report a human-coded validation sample and compare the classifier with pre-trained models (Araci, 2019; Webersinke et al., 2022). For authors and editors, the conditional association between solution-oriented framing and citations is a hypothesis for further work and not a recommendation to change how abstracts are written, since the association may reflect topic or journal effects.

5.3 Limitations

Several limitations qualify the findings. First, the corpus comes from capped exports with the same number of records for each publication year, the sort order and the extraction date are not reported in this article, and the Boolean query strings are not reproduced in full. Only one database and English-language documents were used. The relevance screen is permissive and keyword-based and removed 48% of the records. The manual audit of 300 retained records found that 26.3% are not relevant and that only 57.3% are fully relevant, so that the results of Sections 4.1 to 4.8 describe a broad corpus that includes off-topic records, as the heterogeneous clusters C1 and C2 also show. Only the main analyses were repeated on stricter subsets (Section 4.10), and the records removed by the screen were not audited.

Second, the framing indicator relies on a lexicon whose agreement with human coding is weak on three classes ($\kappa = 0.24$), and the human reference is small: 300 records, 200 of them coded

by one coder and 100 with consensus labels, with no independent inter-coder agreement measure for the final labels. The classifier reproduces the lexicon, the in-sample κ is inflated by training, and the ablation shows that much of the agreement depends on the lexicon words. The lexicon contains the word "risk", which is also the subject of the corpus, and the polarity threshold changes the share of mixed abstracts (from 10.2% to 19.3% between ± 0.01 and ± 0.20), although not the main conclusions. The classification uses a single random split with fixed hyperparameters, and mixed documents are excluded from the classifier. ClimateBERT and FinBERT were used only as frozen embeddings on the 148 binary audit records, without fine-tuning and without application to the full corpus. The cross-database check was made on the initial export and not on the screened corpus, and it rests on rank correlations.

Third, the clustering is weak (silhouette between 0.036 and 0.087), only moderately stable (ARI = 0.55), and agrees only moderately with NMF topics (ARI = 0.36) and weakly with LDA topics (ARI = 0.12). The regression controls for year, open access, team size and, in robustness models, journal and cluster, but not for field or institution, and it does not support causal reading.

5.4 Future research

Five extensions follow from these limitations. Extending the human-annotated set to 600 to 1,000 abstracts with two fully independent coders, followed by fine-tuning ClimateBERT and FinBERT on these labels and applying the best model to the full corpus, would test the validity of the framing indicator and could replace the lexicon. Models of citations with field and institution controls would address the determinants of citations more fully. An audit of a sample of the records removed by the screen and a sensitivity analysis of the queries would establish the content validity of the corpus. A comparison with BERTopic would test the thematic structure further. Finally, co-citation and bibliographic coupling maps would allow a true analysis of the intellectual structure of the field.

6. Conclusion

This study described the bibliometric structure and the framing orientation of 4,164 articles and reviews on climate risk and financial stability published between 2010 and 2025, selected from a Dimensions export of 8,000 records. Its main results concern four questions, and they are descriptive.

On the bibliometric structure (RQ1), citations are strongly concentrated, three citation databases order papers consistently, a relevance screen removed nearly half of the initial export, and about a quarter of the retained records are still not relevant according to the audit. On the reproduction of framing labels (RQ2), a linear SVM trained on lexicon-based labels reproduces them closely, which reflects internal consistency, while a manual audit of 300 records shows weak agreement between the lexicon and human framing. On the evolution of framing (RQ3), the data do not establish a trend, and the proposition of a shift toward solutions is not supported. On thematic clusters (RQ4), eight clusters differ markedly in framing, partly by construction, and solution-oriented abstracts have higher expected citations conditional on controls.

These results suggest targeted recommendations. Evaluators should combine field-normalised and raw citation indicators and report how a corpus was built and screened. Researchers using text classification for framing or sentiment should validate their labels against human coding, as the weak agreement found here ($\kappa = 0.24$ on three classes) shows, and test how much of the performance depends on the lexicon words. The perspectives are testable questions: whether human-coded framing matches lexicon-based framing, whether solution-oriented and risk-oriented papers differ in citations after controlling for age, journal and field, and whether the clusters are stable across methods.

References

- (1). Acharya, V. V., Berner, R., Engle, R., Jung, H., Stroebl, J., Zeng, X., & Zhao, Y. (2023). Climate stress testing. *Annual Review of Financial Economics*, 15, 291–326.
- (2). Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies*, 30(1), 2–47.
- (3). Alogoskoufis, S., Dunz, N., Emambakhsh, T., Hennig, T., Kaijser, M., Kouratzoglou, C., Muñoz, M. A., Parisi, L., & Salleo, C. (2021). ECB's economy-wide climate stress test (ECB Occasional Paper No. 281). European Central Bank.
- (4). Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. arXiv:1908.10063.
- (5). Ardia, D., Bluteau, K., Boudt, K., & Inghelbrecht, K. (2023). Climate change concerns and the performance of green versus brown stocks. *Management Science*, 69(12), 7607–7632.
- (6). Battiston, S., Mandel, A., Monasterolo, I., Schütze, F., & Visentin, G. (2017). A climate stress-test of the financial system. *Nature Climate Change*, 7(4), 283–288.
- (7). Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370.
- (8). Bolton, P., & Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, 142(2), 517–549.
- (9). Bolton, P., & Kacperczyk, M. (2023). Global pricing of carbon-transition risk. *The Journal of Finance*, 78(6), 3677–3754.
- (10). Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. *The Review of Financial Studies*, 22(6), 2201–2238.
- (11). Campiglio, E., Dafermos, Y., Monnin, P., Ryan-Collins, J., Schotten, G., & Tanaka, M. (2018). Climate change challenges for central banks and financial regulators. *Nature Climate Change*, 8, 462–468.
- (12). Carney, M. (2015, September 29). Breaking the tragedy of the horizon: Climate change and financial stability [Speech]. Lloyd's of London. <https://www.bankofengland.co.uk/speech/2015/breaking-the-tragedy-of-the-horizon-climate-change-and-financial-stability>
- (13). Choi, D., Gao, Z., & Jiang, W. (2020). Attention to global warming. *The Review of Financial Studies*, 33(3), 1112–1145.
- (14). Christophers, B. (2017). Climate change and financial instability: Risk disclosure and the problematics of neoliberal governance. *Annals of the American Association of Geographers*, 107(5), 1108–1127.
- (15). Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- (16). Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- (17). Dafermos, Y., Nikolaidi, M., & Galanis, G. (2018). Climate change, financial stability and monetary policy. *Ecological Economics*, 152, 219–234.
- (18). Dietz, S., Bowen, A., Dixon, C., & Gradwell, P. (2016). 'Climate value at risk' of global financial assets. *Nature Climate Change*, 6(7), 676–679.
- (19). Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296.
- (20). DORA. (2013). San Francisco Declaration on Research Assessment. <https://sfdora.org/read/>
- (21). Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebl, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3), 1184–1216.

- (22). Fankhauser, S., Smith, J. B., & Tol, R. S. J. (1999). Weathering climate change: Some simple rules to guide adaptation decisions. *Ecological Economics*, 30(1), 67–78.
- (23). Flammer, C. (2021). Corporate green bonds. *Journal of Financial Economics*, 142(2), 499–516.
- (24). Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(4060), 471–479.
- (25). Giglio, S., Kelly, B., & Stroebe, J. (2021). Climate finance. *Annual Review of Financial Economics*, 13, 15–36.
- (26). Gros, D., Lane, P. R., Langfield, S., Matikainen, S., Pagano, M., Schoenmaker, D., & Suarez, J. (2016). Too late, too sudden: Transition to a low-carbon economy and systemic risk (Reports of the Advisory Scientific Committee No. 6). European Systemic Risk Board.
- (27). Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431.
- (28). Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.
- (29). Hong, H., Li, F. W., & Xu, J. (2019). Climate risks and market efficiency. *Journal of Econometrics*, 208(1), 265–281.
- (30). Hook, D. W., Porter, S. J., & Herzog, C. (2018). Dimensions: Building context for search and evaluation. *Frontiers in Research Metrics and Analytics*, 3, 23.
- (31). Hutchins, B. I., Yuan, X., Anderson, J. M., & Santangelo, G. M. (2016). Relative Citation Ratio (RCR): A new metric that uses citation rates to measure influence at the article level. *PLoS Biology*, 14(9), e1002541.
- (32). IPCC. (2014). Climate change 2014: Impacts, adaptation, and vulnerability. Part B: Regional aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
- (33). IPCC. (2021). Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. <https://doi.org/10.1017/9781009157896>
- (34). Ivanov, I. T., Krutli, M. S., & Watugala, S. W. (2024). Banking on carbon: Corporate lending and cap-and-trade policy. *The Review of Financial Studies*, 37(5), 1640–1684.
- (35). Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *Machine Learning: ECML-98 (Lecture Notes in Computer Science, Vol. 1398, pp. 137–142)*. Springer.
- (36). Krueger, P., Sautner, Z., & Starks, L. T. (2020). The importance of climate risks for institutional investors. *The Review of Financial Studies*, 33(3), 1067–1111.
- (37). Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- (38). Leal Filho, W., Azeiteiro, U., Alves, F., Pace, P., Mifsud, M., Brandli, L., Caeiro, S. S., & Disterheft, A. (2018). Reinvigorating the sustainable development research agenda: The role of the sustainable development goals (SDG). *International Journal of Sustainable Development & World Ecology*, 25(2), 131–142.
- (39). Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35–65.
- (40). Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56–63.
- (41). Milojević, S., & Leydesdorff, L. (2022). Bibliometrics/scientometrics. In *Encyclopedia of big data (pp. 72–75)*. Springer.
- (42). Moed, H. F. (2005). Citation analysis in research evaluation. Springer.
- (43). Monasterolo, I., & De Angelis, L. (2020). Blind to carbon risk? An analysis of stock market reaction to the Paris Agreement. *Ecological Economics*, 170, 106571.

- (44). Moser, S. C. (2014). Communicating adaptation to climate change: The art and science of public engagement when climate change comes home. *Wiley Interdisciplinary Reviews: Climate Change*, 5(3), 337–358.
- (45). Network for Greening the Financial System. (2019). A call for action: Climate change as a source of financial risk. NGFS.
- (46). Network for Greening the Financial System. (2021). NGFS climate scenarios for central banks and supervisors (June 2021). NGFS.
- (47). Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2022). Dissecting green returns. *Journal of Financial Economics*, 146(2), 403–424.
- (48). Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- (49). Price, D. J. de S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5), 292–306.
- (50). Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv:2205.01833.
- (51). Reboredo, J. C., & Ugolini, A. (2022). Climate transition risk, profitability and stock prices. *International Review of Financial Analysis*, 83, 102271.
- (52). Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- (53). Sautner, Z., van Lent, L., Vilkov, G., & Zhang, R. (2023). Firm-level climate change exposure. *The Journal of Finance*, 78(3), 1449–1498.
- (54). Sauzet, M., & Zerbib, O. D. (2022). When green investors are green consumers. SSRN Working Paper No. 4098124.
- (55). Stern, N. (2008). The economics of climate change. *American Economic Review*, 98(2), 1–37.
- (56). Task Force on Climate-related Financial Disclosures. (2017). Final report: Recommendations of the Task Force on Climate-related Financial Disclosures. Financial Stability Board.
- (57). Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.
- (58). Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5), 988–999.
- (59). Vinkers, C. H., Tjldink, J. K., & Otte, W. M. (2015). Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: Retrospective analysis. *BMJ*, 351, h6467.
- (60). Violle, A. (2025). A climate for investors: Climate scenarios in the Network for Greening the Financial System. *Valuation Studies*, 12, 258–282.
- (61). Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20–41.
- (62). Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391.
- (63). Webersinke, N., Kraus, M., Bingler, J. A., & Leippold, M. (2022). ClimateBERT: A pretrained language model for climate-related text. arXiv:2110.12010.
- (64). Weitzman, M. L. (2009). On modeling and interpreting the economics of catastrophic climate change. *The Review of Economics and Statistics*, 91(1), 1–19.
- (65). Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429–472.