Journal of Financial Stability 84 (2026) 101549

Contents lists available at ScienceDirect

Journal of Financial Stability

journal homepage: www.elsevier.com/locate/jfs

Climate risk news and banking industry: A natural language processing approach[☆]

Babak Naysary [a,*,1], Ali Edisen [a], Amine Tarazi [b,c]

[a] Birmingham City University, Business School, 15 Bartholomew Row, Birmingham B5 5JU, United Kingdom
[b] Université de Limoges, LAPE, 5 rue Félix Eboué, Cedex 1, Limoges 87031, France
[c] Institut Universitaire de France (IUF), 1 rue Descartes, Cedex 05, Paris 75231, France

ARTICLE INFO

Keywords:
Banking
Climate risk
NLP
Text mining
Sentiment analysis
Topic modeling

ABSTRACT

This study analyzes the evolution of climate-risk discourse in banking using 4887 news articles (2008–2024) collected from ProQuest. We apply Natural Language Processing and add two novel layers: (i) an event-alignment analysis that links coverage dynamics to dated policy and supervisory milestones, and (ii) a discourse-network analysis connecting banks and regulators. We document a marked post-2020 shift, with ESG emerging as the dominant framing (7860 mentions) alongside persistent geographic asymmetries (U.S.-led coverage) and uneven sectoral engagement (Risk Management highest salience; Fintech lowest). Sentiment skews positive (≈4000 positive vs. ≈1500 negative), and topic modeling identifies eight stable thematic clusters spanning operations, ratings, ESG assessment, disclosures, and market instruments. Event alignment shows media attention is typically anticipatory (median peak two months before an anchor), with COP26 producing a sustained level shift (+100% within a ± 6-month window) and the Bank of England's CBES results generating the largest single spike (210 articles), whereas some 2022 rule-making announcements (e.g., SEC climate-disclosure proposal) exhibit sharper but less durable attention. The discourse network centers on two regulatory hubs (the Federal Reserve and the ECB) with key banks (e.g., Citigroup, JPMorgan, UBS) bridging into supervisory narratives. Collectively, the findings show climate risk becoming embedded in core banking practice while revealing structural, regional, and functional asymmetries that matter for policy design and implementation.

1. Introduction

With US$1.6 trillion in banking exposures to high environmental risk sectors across major economies (Nieto, 2019), climate risk has evolved from a peripheral environmental concern to a systemic financial stability issue. Yet despite this massive exposure, our understanding of how this transformation actually unfolded remains fragmented. The urgency of addressing this knowledge gap became apparent when the Federal Reserve (2024) pilot exercise revealed that even the six largest U.S. banks struggled to reliably assess their climate risk exposures.[2] This failure at the industry's highest level highlights fundamental disconnects in how climate risk is conceptualized and implemented across the banking sector. This study addresses this critical gap by systematically analyzing the evolution of climate risk discourse in banking revealing not just how banks' approaches transformed over time but trying to address why such significant variations in climate risk integration

☆ The authors are grateful to the anonymous reviewers and the Guest Editor for their constructive comments and suggestions, which have substantially improved the quality of this paper.

* Corresponding author.
E-mail addresses: babak.naysary@bcu.ac.uk, bnaysary@hotmail.com (B. Naysary), ali.edisen@mail.bcu.ac.uk (A. Edisen), amine.tarazi@unilim.fr (A. Tarazi).
1 ORCiD: 0000-0003-0464-7402

2 The Federal Reserve, (2024) Pilot Climate Scenario Analysis Exercise showed that participants (Bank of America Corporation; Citigroup Inc; The Goldman Sachs Group, Inc.; JPMorgan Chase & Co.; Morgan Stanley; and Wells Fargo & Company) are facing significant data and modeling challenges and disparities in estimating climate-related financial risks. The resulting "variation in the level of sophistication of climate risk modelling approaches" makes it impossible to "compare risks across the industry," as each institution effectively operates with its own climate risk definition and measurement system (KPMG, 2024). These conceptual and methodological disparities persist globally, with European assessments finding that approaches "still lack methodological sophistication" and "more than half of the banks under the ECB's supervision not implementing the practices effectively'' (ECB, 2024), while the IMF identifies "persistent large gaps in climate data and disclosures" that prevent meaningful cross-institutional comparison (IMF, 2023).

https://doi.org/10.1016/j.jfs.2026.101549
Received 21 January 2025; Received in revised form 27 April 2026; Accepted 6 May 2026
Available online 9 May 2026

persist across institutions today.

The evolution of climate risk in banking represents a fundamental shift in how financial institutions perceive and manage environmental challenges (Battiston et al., 2021; Zhai et al., 2024). It has become a central element in banking strategy and risk management, with recent studies revealing that banks' climate risk exposure through lending significantly impacts their tail risks and systemic risk contribution (Conlon et al., 2024). This transformation has been driven by growing recognition of two distinct but interconnected challenges: physical risks, encompassing direct environmental impacts on assets and operations, and transition risks, associated with the global shift toward a low-carbon economy (Feridun and Güngör, 2020; Monasterolo, 2020). The significance of these risks is shown by research reporting that firms in locations with higher exposure to climate risk pay significantly higher spreads on bank loans (Javadi and Masum, 2021), and banks are increasingly adjusting their lending practices based on firms' carbon emissions (Altavilla et al., 2024). The institutionalization of these concerns has been accelerated by initiatives such as the Task Force on Climate-Related Financial Disclosures (TCFD) and the Glasgow Financial Alliance for Net Zero (GFANZ), though recent research indicates that current climate models and detailed databases remain inadequate for fully estimating climate risks in the banking industry (Cardenas, 2024).

The transformation's significance extends beyond mere environmental considerations. Research shows that banks with significant climate risk exposure through lending face increased tail risks and systemic risk contribution (Conlon et al., 2024), while firms in climate-vulnerable locations encounter substantially higher loan spreads, particularly those with poor credit ratings (Javadi and Masum, 2021). These findings suggest that climate risk has fundamentally changed the risk-return dynamics of banking operations, however, the evolution of how banks have collectively recognized, discussed, and responded to these changes remains mainly unexplored.

Moreover, the existing literature's focus on individual institutional responses or isolated aspects of climate risk management has created what might be termed "institutional myopia". While studies document how specific banks approach climate risk assessment (Liu et al., 2024; Pasquini, 2020) or analyze particular risk dimensions such as credit risk impacts (Altavilla et al., 2024) and capital structure adjustments (Ginglinger and Moreau, 2023), these approaches fail to capture the collective evolution of banking sector discourse and practice. This limitation is particularly significant because of substantial variations in climate risk integration across different banking sectors and across different geographical regions, which may reflect deeper structural differences (Yang et al., 2023; Zhang et al., 2024).

From a methodological perspective, most existing studies rely on survey data, regulatory filings, or case studies that provide valuable but limited perspectives on how the banking sector's collective understanding and approach to climate risk has developed over time (Battiston et al., 2021; Monasterolo, 2020). These approaches, while offering important insights into individual institutional responses, are limited in their ability to show temporal patterns, thematic evolution, and the interplay between market-driven innovation and regulatory development.

To address these critical gaps, this study provides the first comprehensive temporal analysis of climate risk discourse evolution in banking through systematic examination of 4887 English-language news articles collected from ProQuest spanning from 2008 to 2024. Our methodological approach utilizes advanced Natural Language Processing techniques (including term frequency analysis, sentiment analysis, topic modeling using Latent Dirichlet Allocation, and Named Entity Recognition) to extract meaningful patterns from nearly two decades of banking sector climate risk discourse. Our findings advance the literature on climate risk and financial institutions in several important ways.

First, we provide the first longitudinal and systematic analysis of climate risk discourse in the global banking sector using large-scale unstructured textual data. While prior research has examined climate risk in banking primarily through regulatory filings, surveys, or case-specific analyses (Battiston et al., 2021; Liu et al., 2024; Yang et al., 2023), such approaches often lack the temporal coverage and industry-wide view, necessary to capture evolving patterns. We identify structural shifts in the salience, framing, and thematic composition of climate risk discourse, including the post-2020 surge in ESG-oriented narratives. This longitudinal perspective addresses calls in recent reviews (Cardenas, 2024; Zhai et al., 2024) for more comprehensive empirical evidence on sector-wide climate risk integration.

Second, we advance understanding of the drivers behind climate risk discourse evolution through event alignment analysis that systematically links discourse patterns to specific regulatory milestones and market developments. By examining discourse responses before and after major events such as the Paris Agreement (2015), TCFD launch (2017), and COP26/GFANZ commitments (2021), we reveal whether banks' climate risk integration has been strategic or reactive to external pressures. This approach provides crucial insights into which regulatory interventions generate sustained engagement versus temporary responses, addressing ongoing debates about the primary drivers of climate risk integration in banking (Feridun and Güngör, 2020; Roncoroni et al., 2021).

Third, we contribute novel insights into the institutional dynamics of climate risk discourse through network analysis mapping relationships between major banking institutions, regulators, and key climate risk concepts. This analysis reveals which banks serve as discourse leaders versus followers, identifying the pathways through which climate risk innovations and practices spread across the industry.

Fourth, our study offers comprehensive sectoral and geographical insights into how climate risk integration varies within the banking industry. While prior work has documented cross-country differences in climate risk exposure and response (Nieto, 2019; Zhang et al., 2024), there has been no systematic, comparative mapping of these variations across both banking sub-sectors and geographical regions. Our findings reveal structural asymmetries (such as the dominance of U.S.-led discourse and the relatively low visibility of climate risk in Fintech) while demonstrating how different banking functions exhibit varying levels of climate risk integration, from Risk Management's leading position to emerging sectors' gradual adoption.

Fifth, we advance the literature through multi-dimensional thematic and sentiment analysis that reveals the evolution from risk-avoidance to opportunity-oriented climate risk framing. Our combined use of Latent Dirichlet Allocation (LDA) topic modeling, sentiment classification, and co-occurrence analysis demonstrates how ESG has emerged as the dominant integrative framework for climate risk in banking. Our empirical evidence showing ESG-related terms appearing over four times more frequently than direct "climate risk" references substantiates claims about ESG's centrality to sustainable finance discourse (Krueger et al., 2020; Milkau, 2022), while mapping co-occurrence patterns with terms such as "credit rating" and "risk management" provides operational-level insights into how climate considerations are embedded in banking decisions.

Finally, we make a methodological contribution by demonstrating how advanced NLP techniques can systematically analyze financial news coverage to complement traditional quantitative and qualitative methods in climate finance research. While NLP has been increasingly used in market prediction and sentiment extraction (Du et al., 2025; Lo and Singh, 2023), its application to climate finance research remains nascent. Our integrated analytical framework, synthesizing evidence across temporal, geographical, sectoral, thematic, sentiment, event-driven, and network dimensions, is adaptable for future research examining other systemic risks in the financial sector.

The remainder of this paper is structured as follows: 2 provides a comprehensive review of related literature on the implications of climate risk in financial systems as a whole and banking industry in particular, and also the development of NLP methods in analyzing financial news. 3 presents our data collection methodology and analytical framework. 4 provides the results which include examining the temporal and geographical distribution of climate risk discourse, analyzing the key themes and topics emerging from our analysis, results of our sentiment analysis and topic modeling results, providing insights into how the banking industry's approach to climate risk has evolved. 5 includes the final discussion of our findings and concluding remarks.

2. Literature review

2.1. Climate risk and financial institutions

Chenet (2021) provided a comprehensive framework for analyzing how climate change translates into financial risks, distinguishing between physical risks arising from climate-related events and transition risks associated with policy and technological changes toward a low-carbon economy. This conceptual foundation was further developed by Monasterolo (2020), who emphasized the systemic nature of climate risks and their potential to create cascading effects throughout the financial system. Building on these foundations, Battiston et al. (2021) highlighted critical methodological gaps in integrating climate considerations into financial risk assessment, arguing that traditional risk management frameworks are inadequate for capturing the unique characteristics of climate risks. The theoretical literature has also addressed the challenge of quantifying climate risks. Mandel et al. (2021) developed sophisticated modeling approaches for flood risks' impact on financial stability, while Zhou et al. (2023) examined how various factors such as income levels and market diversification can mitigate negative climate risk impacts. These studies demonstrate the complexity of translating climate science into financial risk metrics, a challenge that remains central to current implementation difficulties.

The empirical literature has provided increasing evidence of climate risk impacts on banking operations. Conlon et al. (2024) demonstrate that banks' climate risk exposure through lending significantly increases their tail risks and systemic risk contribution, providing quantitative evidence that climate considerations are not merely environmental concerns but fundamental financial risks. This finding is complemented by Javadi and Masum (2021), who show that firms in climate-vulnerable locations face significantly higher loan spreads, particularly those with poor credit ratings, indicating that climate risk is already being priced into banking relationships. Recent empirical work has extended these findings to demonstrate active responses by banks. Altavilla et al. (2024) provide evidence that banks are increasingly adjusting their lending practices based on firms' carbon emissions, charging higher rates to high-emission firms while offering preferential rates to those committed to emission reductions. This evidence suggests that climate risk considerations have moved beyond risk assessment into active portfolio management and pricing strategies. Geographical variations in climate risk impacts have also been documented extensively. Liu et al. (2024) analyze Chinese commercial banks, showing that climate change shocks significantly increase credit risk, particularly affecting small-scale and rural banks. Liu et al. (2023) document regional disparities in climate transition risk within China, with higher levels in eastern regions and lower levels in central-western areas. These studies highlight how climate risk impacts vary not only by institution but also by geography and market structure.

The literature documents significant variation in how financial institutions have responded to climate risks. Krueger et al. (2020) find that institutional investors increasingly recognize climate risks' financial implications, preferring risk management and engagement strategies over simple divestment. However, implementation of these strategies faces substantial challenges, with Taylor (2023) documenting how actuaries struggle to incorporate climate-related risks into professional frameworks due to inadequate risk management tools. European research provides insights into regional approaches to climate risk management. Pasquini (2020) examines Italian banks' increasing focus on climate risk management, highlighting the need for investments in databases and methodologies. Nocoń (2024) documents the Polish banking sector's integration of climate risk management, while broader European assessments reveal persistent challenges in implementation sophistication. The literature also highlights the role of regulatory frameworks in shaping institutional responses. Feridun and Güngör (2020) review emerging regulatory and supervisory practices, emphasizing the importance of board-level attention and scenario analysis. Lamperti et al. (2021) propose specific policy measures, including green Basel-type capital requirements and carbon-risk adjustments, demonstrating how regulatory approaches can influence bank behavior while maintaining financial stability. The regulatory literature reveals a complex evolution in supervisory approaches to climate risk. Campiglio et al. (2018) emphasize the need for central banks and financial regulators to develop comprehensive analytical frameworks for assessing climate change impacts. This call for regulatory development has been followed by various initiatives, though implementation remains challenging. Roncoroni et al. (2021) demonstrate how climate policy shocks can propagate through networks of banks and investment funds, highlighting the systemic nature of climate risks and the need for coordinated regulatory responses. The literature suggests that regulatory approaches must account for both direct institutional impacts and network effects that can amplify risks across the financial system.

2.2. Natural language processing in finance

The application of NLP in financial analysis has gone through dramatic evolution, creating new possibilities for understanding complex financial phenomena through textual data analysis. Lo and Singh (2023) provide a comprehensive overview of how NLP in finance has evolved from early rule-based systems to sophisticated deep learning approaches. This evolution has been marked by increasing capability to process and understand financial language, with modern approaches capable of extracting insights from complex financial texts. The development of specialized financial language models, such as FinGPT by Yang et al. (2023), demonstrates the growing sophistication of finance-specific NLP applications. The methodological literature emphasizes the importance of domain-specific adaptation in financial NLP. Upreti et al. (2019) demonstrate how financial news analytics can be enhanced by integrating domain expertise with linguistic information, while Achitouv et al. (2023) show how NLP can effectively match semantic patterns in financial regulations without requiring extensive datasets. These studies highlight the need for specialized approaches when applying NLP to financial contexts.

A substantial body of research focuses on sentiment analysis applications in financial markets. Takale (2024) shows how modern NLP techniques have improved the accuracy and efficiency of financial sentiment analysis, leading to more reliable market forecasts. Yinhao (2024) provides evidence that NLP technologies can effectively analyze the impact of financial news on stock market movements, particularly for short-term predictions.

Early work by Alvim et al. (2010) demonstrated the superiority of machine learning approaches over traditional methods in financial sentiment classification. This foundational work has evolved into more sophisticated frameworks that combine multiple analytical techniques to extract insights from financial communications. The application of text mining techniques to financial data has revealed new possibilities for understanding market dynamics and institutional behavior. Farimani et al. (2022) highlight the potential of combining text mining with deep learning methods for improved market prediction, while noting computational and data quality challenges that remain significant barriers to implementation. Xiao et al. (2024) demonstrate how NLP technology can provide insights into market dynamics while improving customer service and investment strategy innovation. This work illustrates the practical applications of NLP beyond simple text analysis to strategic business insights.

Recent research has emphasized making financial NLP more accessible and inclusive. Ghosh and Naskar (2023) demonstrate how NLP tools can enhance the understandability of financial texts, contributing to improved financial literacy. This democratization extends to language accessibility, with works like IndicFinNLP (Ghosh et al., 2024) addressing the need for financial NLP resources in non-English languages. These developments align with broader trends toward making financial analysis more accessible to diverse user groups, though most climate risk research remains focused on English-language sources, representing a significant limitation in current approaches.

The literature increasingly emphasizes the complementary nature of NLP and traditional financial analysis methods. Oyewole et al. (2024) document how NLP integration in financial reporting improves both accuracy and efficiency, while noting implementation challenges and regulatory considerations. This integration represents a significant methodological advance, allowing researchers to combine quantitative financial data with qualitative textual insights. Uğur et al. (2022) demonstrate how modern NLP techniques, particularly deep learning architectures, can effectively analyze online news channels for investment opportunities and risks. This work illustrates how NLP can extend traditional financial analysis by incorporating previously unavailable information sources.

2.3. Literature gaps

Despite the extensive literature on climate risk in banking and the growing sophistication of NLP methods in financial analysis, several critical gaps persist that limit our understanding of how climate risk integration has evolved in the banking sector.

The existing literature predominantly focuses on current practices or isolated time periods, lacking comprehensive temporal analysis of how climate risk discourse and practice have evolved over time (e.g., Battiston et al., 2021; Conlon et al., 2024; Krueger et al., 2020; Liu et al., 2024; Monasterolo, 2020; Pasquini, 2020).[3] For instance, while studies like Battiston et al. (2021); Monasterolo (2020) provide snapshots of climate risk challenges, they do not capture the dynamic evolution of how banks have collectively approached these challenges. This temporal gap is particularly significant given evidence from recent regulatory assessments (ECB, 2024; Federal Reserve, 2024) that implementation varies dramatically across institutions, suggesting different evolutionary paths that remain unexplored.

Similarly, while the NLP literature demonstrates sophisticated analytical capabilities (Du et al., 2025; Lo and Singh, 2023), its application to tracking the evolution of climate risk discourse in banking remains basic. Most NLP studies in finance focus on market prediction or sentiment analysis for trading purposes, rather than understanding how institutional approaches to complex risks evolve over time (e.g., Alvim et al., 2010; Farimani et al., 2022; Takale, 2024; Uğur et al., 2022; Yinhao, 2024).[4] This methodological gap is particularly significant for climate risk research, where much of the relevant information exists in unstructured formats (regulatory communications, industry reports, news coverage) that are not captured by traditional quantitative methods.

The literature reveals a significant gap in systematic analysis of how different banking sectors and institutions approach climate risk integration. While studies document variations in climate risk response (Liu et al., 2023; Zhang et al., 2024), there has been no comprehensive mapping of how these variations manifest across different banking functions (risk management, commercial banking, investment banking, etc.) or how they have evolved over time. This gap is particularly notable given evidence from recent assessments that banks operate with fundamentally different climate risk definitions and measurement systems (KPMG, 2024). Understanding why these differences emerged requires systematic analysis of how different institutional contexts shaped climate risk integration, which existing studies have not provided.

3. Data and method

3.1. Data acquisition

The dataset comprises English-language news articles obtained from ProQuest, a comprehensive database providing extensive coverage of global industry-specific news.[5] ProQuest was selected for its broad coverage of financial media sources and its robust indexing system, which enables precise targeting of banking-related content.

Following established practices in financial news analysis research (Nach, 2025), we developed a structured search query designed to capture the full spectrum of climate risk discourse in banking:

((climate risk) OR (environmental risk) OR carbon OR sustainability OR

esg OR (green finance)) AND (transition risk) AND (physical risk) AND

(climate change) AND emission AND mainsubject.Exact("Banking industry" OR "Banking law")

This query incorporates multiple dimensions of climate risk terminology while ensuring relevance to banking through the mainsubject filter. The query yielded 6991 potentially relevant articles published between January 2008 and December 2024.

The ProQuest database query yielded 6991 relevant news articles. Due to the platform's technical limitation restricting exports to 500 articles per file, we implemented a systematic extraction process across fourteen sequential batches, ranging from articles 1–500 through 6501–6991. Each export utilized ProQuest's comprehensive "full text" feature to capture complete article metadata, including citations, abstracts, full text content, and subject classifications. The separate files

[3] Further studies exemplify this temporal limitation. Curcio et al. (2023) analyze U.S. banks' and insurers' responses to billion-dollar climate disasters but focus on event-specific impacts rather than tracking discourse evolution over time. Similarly, Li and Pan (2022) examine climate risk's effect on bank performance using cross-sectional data without capturing how banking sector understanding has developed longitudinally. Reghezza et al. (2022) document European banks' lending responses following the Paris Agreement but examine a specific post-policy period rather than the broader temporal evolution of climate risk integration.

[4] This market prediction orientation is evident in other recent works as well. Fatouros et al. (2023) deploy ChatGPT for financial sentiment analysis with explicit focus on forex market returns and trading decisions. Leippold (2023) examines GPT-3's performance in sentiment classification specifically to assess its utility for market-focused applications, while Wu et al. (2023) develop BloombergGPT, a large language model explicitly designed for financial market prediction tasks. These studies demonstrate NLP's sophisticated capabilities but remain oriented toward short-term market applications rather than understanding institutional evolution over time.

[5] Our corpus was drawn from the News feed in ProQuest, whose publisher is recorded as Dow Jones & Company Inc. However, the news items originate from multiple agencies and sources, including Reuters, Morning Risk Report (Wall Street Journal), GlobeNewswire, Business Wire, PR Newswire, AM Best, DBRS Morningstar, Moody’s Investors Service, Fitch Ratings, and S&P Global Ratings. Because originator cues appear in unstructured, non-standardised places and are not consistently present in all records, a robust distributional breakdown is not reliably computable from the export.

were subsequently consolidated using Python's Pandas library, resulting in a unified 86 MB dataset saved as a.csv file where individual articles are distinctly separated by standardized delimiters.

Individual articles within the consolidated dataset are separated by "____" delimiters, with each article containing structured fields including Document ID, Title, Full text, Publication year, Publication date, and Subject classifications. We developed a systematic parsing algorithm implemented in Python to extract these structured attributes from the raw text files using the Pandas library for efficient data manipulation and the pathlib module for robust file handling.

The extraction process involved multiple stages of text preprocessing to ensure data consistency and accuracy. Text normalization procedures standardized various text formats while addressing case inconsistencies and eliminating characters that could interfere with subsequent analysis. Field parsing utilized regular expression patterns implemented through Python's re module to identify and extract specific metadata fields from each article. Date standardization converted various date formats, such as "Jan 12, 2022" and "January 12, 2022", to ISO format (YYYY-MM-DD) for temporal analysis using Python's datetime library.

To ensure dataset quality and relevance, we implemented a multi-stage filtering process addressing concerns about duplicate content and misclassified articles. Our quality control algorithm identifies and removes several categories of problematic content through systematic validation procedures.

Document type filtering excludes non-news content including advertisements, corrections, errata, public notices, classified ads, market data tables, letters to the editor, and stock quotes. Content length validation removes ultra-short articles containing fewer than 80 words, which typically represent brief notices rather than substantive news coverage. Subject relevance verification implements strict subject filtering requiring explicit mention of "bank" in subject classifications, aligning with our mainsubject query filter. Structural validation ensures all retained articles contain essential fields including Document ID and Title, along with valid publication years.

The deduplication process eliminates redundant content through a comprehensive three-stage approach implemented using Pandas' data manipulation capabilities. The first stage removes articles sharing identical ProQuest Document IDs, eliminating one duplicate from our dataset. The second stage addresses content hash deduplication by eliminating articles with identical full text content using MD5 hash comparison through Python's hashlib library, accounting for multi-source reposts of the same article and removing 34 duplicates. The third stage identifies near-duplicates using normalized combinations of title, publication date, and publisher, removing an additional 385 duplicate entries.

After quality control procedures, our final dataset contains 4887 articles, representing a 70% retention rate from the original 6991 articles. The filtering process removed 1684 articles, with the majority eliminated due to insufficient content length (1592 articles, 94.5%), followed by missing essential metadata (53 articles, 3.1%), correction notices (19 articles, 1.1%), calendar items (19 articles, 1.1%), and off-topic subject matter (1 article, 0.1%). This systematic approach to data cleaning ensures that our analysis focuses on substantive, relevant news coverage while maintaining the integrity and representativeness of the climate risk discourse in banking.

3.2. Term frequency and TF-IDF analysis

At this stage after lemmatizing each article and removing generic stop-words, every document d is represented by a sparse vector of weighted terms. The weight begins with the normalized term frequency:

$$TF_{d,t} = \frac{f_{d,t}}{\sum_k f_{d,k}} \tag{1}$$

where $f_{d,t}$ represents the raw count of term t in document d, and $\sum_k f_{d,k}$ denotes the total document length (i.e., the sum of all term frequencies in document d). To reduce the influence of terms that appear frequently across the corpus, this normalized frequency is multiplied by the inverse document frequency:

$$IDF_t = \log\left(\frac{N}{n_t}\right) \tag{2}$$

where N represents the total corpus size and n_t indicates the number of documents containing term t. The final TF-IDF weight is computed as:

$$w_{d,t} = TF_{d,t} \times IDF_t \tag{3}$$

This formulation highlights terms that are both frequent within a specific document and relatively rare across the entire corpus (Gentzkow et al., 2019). We implemented this using scikit-learn’s TfidfVectorizer, configured to extract 5000 unigrams and bigrams, consistent with our TopicAnalyzer implementation.

TF-IDF was selected over contextual sentence embeddings for several practical reasons. First, TF-IDF yields weights that explicitly separate term frequency and inverse document frequency, providing direct interpretability and auditability (Manning, 2008). Second, the sparseness of the TF-IDF feature matrix allows efficient storage and computation, enabling the processing of large corpora on standard hardware without the resource demands or stochastic variability of transformer-based models requiring GPU acceleration (Gentzkow et al., 2019). Third, empirical tests demonstrated that increasing the number of TF-IDF features beyond 5000 did not enhance clustering performance or topic coherence, while the adoption of transformer embeddings incurred substantially higher computational cost without proportional improvement in methodological outcomes, thus rendering TF-IDF the optimal balance of interpretability, efficiency, and effectiveness for this study (Vaswani et al., 2017). This approach supports the identification of prominent banks, countries, and sectoral references through term salience and subsequent aggregation. The objective of this method is to identify and weight the most salient terms and phrases in the news articles, highlighting key concepts in climate risk discourse while accounting for their relative importance across the corpus. Results from TF-IDF analysis support the results presented across multiple subsections of 4 such as 4.2 Insights on countries, 4.3 Most mentioned banks, 4.4 Salient financial sectors, and 4.5 Insights on investment and regulation related discourse.

3.3. Sentiment analysis

Each article undergoes sentiment classification using DistilRoBERTa-FinNews, a domain-specific transformer model built upon RoBERTa (Liu et al., 2019) and DistilBERT architectures *(*Sanh et al., 2019*)*. The model converts input sequences of WordPiece tokens into 768-dimensional contextual vectors, which are subsequently processed through a fully connected classification layer. The output logits z are transformed via softmax normalization:

$$\widehat{p_c} = \frac{\exp(z_c)}{\sum_{j=1}^{3}\exp(z_j)} \tag{4}$$

yielding class probabilities for positive, neutral, and negative sentiment categories. Final classification selects the highest-probability class while recording the associated probability as a confidence measure.

This transformer-based approach demonstrates superior performance compared to traditional methods in resolving context-dependent financial phrases (Liu et al., 2019; Sanh et al., 2019). For instance, "risk reduction" typically conveys positive sentiment, while "risk exposure" indicates negative sentiment (distinctions that lexicon-based approaches often fail to capture) (Du et al., 2024). Model validation employed stratified five-fold cross-validation on a manually annotated subset of 1000 articles (Du et al., 2024), achieving a macro F1-score of 0.88. This substantially outperformed baseline approaches, including TF-IDF features with Support Vector Machines (F1 = 0.71) and the Loughran-McDonald financial sentiment lexicon (F1 = 0.63) (Du et al., 2024). The objective of sentiment analysis is to classify the overall tone of climate risk coverage and track how sentiment has evolved over the study period, revealing whether discourse frames climate risk predominantly as a threat, opportunity, or neutral consideration. Results from sentiment classification are presented in 4.8 Sentiment analysis, which examines both the aggregate distribution of sentiments and their temporal evolution from 2009 to 2024.

3.4. Topic modeling

Topic discovery employs Latent Dirichlet Allocation (LDA) using mean-field variational Bayes optimization, implemented through scikit-learn's LatentDirichletAllocation class. The algorithm maximizes an evidence lower bound approximation to the posterior distributions over document-topic and topic-word relationships. Conceptually, these posteriors follow Dirichlet distributions with updated parameters:

$$\theta_d|z \sim \text{Dirichlet}(\alpha + n_{d,\cdot}) \tag{5}$$

$$\phi_k|z \sim \text{Dirichlet}(\beta + n_{\cdot,k}) \tag{6}$$

where $n_{d,k}$ represents the count of words in document d assigned to topic k, $n_{d,\cdot}$ denotes the total word count for document d, and $n_{\cdot,k}$ indicates the corpus-wide word count for topic k.

Model hyperparameters were selected through systematic optimization of topic coherence scores. Setting $K = 8$ topics, $\alpha = 0.1$, and $\beta = 0.01$ maximized the C_v coherence measure at 0.53, indicating moderate-to-good topic interpretability according to established benchmarks (Blei et al., 2017). The coherence score reflects the degree of semantic similarity among high-probability words within each topic, with values above 0.4 generally considered acceptable for practical applications.

LDA was preferred over alternative approaches including non-negative matrix factorization and BERTopic for several reasons (Blei et al., 2017; Blei et al., 2003). LDA's probabilistic framework enables individual articles to leverage information from the entire corpus through Dirichlet priors, producing stable topic assignments even for infrequent themes (Blei et al., 2003). Additionally, LDA maintains computational efficiency, training approximately ten times faster on standard CPU hardware compared to transformer-based alternatives while preserving full methodological transparency (Blei et al., 2017). The objective of topic modeling using LDA is to discover the latent thematic structure underlying climate risk discourse in banking by identifying probabilistic distributions of co-occurring terms. The topics identified through LDA guide the thematic interpretation presented in 4.9 Topic modeling.

3.5. Document clustering and visualization

Semantic relationships among documents are visualized through dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP). The algorithm constructs a k-nearest-neighbor graph in the original TF-IDF feature space, builds an equivalent fuzzy simplicial complex in the target low-dimensional space, and minimizes the cross-entropy between these topological representations. We configured UMAP with $n_{neighbors} = 15$ and $min_{dist} = 0.1$ following established practices for document visualization (Becht et al., 2019).

The resulting two and three-dimensional coordinates undergo clustering via k-means, which minimizes the within-cluster sum of squares:

$$\sum_{i=1}^{N} \| x_i - \mu_{c(i)} \|^2 \tag{7}$$

where x_i represents the UMAP coordinates for document i, $\mu_{c(i)}$ denotes the centroid of cluster $c(i)$, and $c(i)$ indicates the cluster assignment for document i. The optimal cluster count $k = 8$ was determined through elbow analysis combined with silhouette score evaluation, achieving a silhouette coefficient of 0.41. This indicates moderate clustering quality, as silhouette scores range from -1 to $+1$, with values above 0.3 typically considered acceptable for exploratory analysis.

UMAP was selected over alternative dimensionality reduction techniques due to its superior preservation of both local neighborhood structures and global topological features (Becht et al., 2019). While t-SNE effectively maintains local relationships, it tends to distort large-scale distances between distinct document clusters (Maaten and Hinton, 2008). Principal Component Analysis, conversely, assumes linear relationships that inadequately capture the non-linear semantic structures inherent in textual data (Jolliffe and Cadima, 2016). Furthermore, k-means clustering scales linearly with sample size, making it computationally feasible for the full document corpus, whereas hierarchical clustering methods exhibit quadratic complexity and proved impractical at this scale (Manning, 2008). The objective of document clustering and visualization is to represent the topic structure identified by LDA in an interpretable visual form. Together, these methods enable visual exploration of how the eight thematic domains relate to one another in the corpus. Results are presented in 4.9 Topic modeling including the visualization (Fig. 11).

3.6. Aspect and theme tracking

Temporal analysis of specific climate risk themes employs a curated dictionary of fifteen domain-relevant phrases, such as "net zero," "transition risk," "physical risk," "carbon pricing," and "ESG integration." These terms were selected based on regulatory frameworks, academic literature, and industry practice guidelines to ensure comprehensive coverage of climate risk discourse dimensions.

Monthly occurrence counts for each theme form the matrix $X_{t,m}$, where t indexes themes and m represents monthly time periods. Sentiment confidence scores from the transformer model are similarly aggregated over monthly intervals, providing temporal sentiment trajectories for each thematic category. Pairwise correlations among themes employ Pearson correlation coefficients:

$$\rho_{ij} = \frac{\sum_t (X_{t,i} - \overline{X}_i)(X_{t,j} - \overline{X}_j)}{\sqrt{\sum_t (X_{t,i} - \overline{X}_i)^2}\sqrt{\sum_t (X_{t,j} - \overline{X}_j)^2}} \tag{8}$$

where $\overline{X}_i$ represents the temporal mean for theme i.

This keyword-based approach was preferred over dynamic topic modeling for several methodological advantages. Fixed vocabularies provide complete auditability, enabling supervisors and practitioners to identify precisely which term combinations activate specific thematic categories. Static dictionaries also avoid the topic drift phenomena that commonly affect dynamic models over extended temporal panels, while completing analysis in seconds rather than hours. Despite this computational efficiency, the approach integrates seamlessly with sentiment scores and maintains consistency with the broader analytical framework (Hosseiny Marani and Baumer, 2023).

We employ spaCy's transformer-based Named Entity Recognition (NER) pipeline (en_core_web_trf) to extract named entities from the original article text, preserving capitalization and punctuation critical for accurate entity boundary detection. To enhance recognition of domain-specific terminology, we augment the pre-trained model with a custom EntityRuler containing patterns for major financial institutions ("JPMorgan," "Bank of America," "Deutsche Bank"), climate risk concepts ("net zero," "transition risk"), and financial sectors ("investment banking," "asset management"). This hybrid approach combines the contextual understanding of transformer models with the precision of rule-based matching for specialized financial vocabulary (Mohit, 2014). The objective of aspect and theme tracking is to examine the temporal evolution of key climate risk concepts in banking-related news and to assess how the prominence of specific themes changes over time. This approach enables analysis of changes in thematic emphasis and the relationships among concepts as reflected in their co-occurrence patterns. The outcomes of this analysis are presented in 4.6, which reports overall climate risk theme frequencies, and 4.7, which examines temporal co-occurrence patterns among selected concepts.

3.7. Event alignment of news coverage

We quantify the relationship between major climate–finance milestones and media attention by aligning monthly news volumes to dated anchor events (Lochner et al., 2024). Starting from the cleaned corpus, we parse the publication date for each article, and aggregate to the first day of each month to obtain a univariate time series of article counts $\{y_t\}$ at monthly frequency. For each event e with calendar date t_e, we define a symmetric window of $\pm w$ months (default $w = 6$), treating months strictly before t_e as the pre-period and the event month and subsequent w months as the post-period. Let $\mathscr{T}_e^{\text{pre}} = \{t_e - w, \ldots, t_e - 1\}$ and $\mathscr{T}_e^{\text{post}} = \{t_e, \ldots, t_e + w\}$, with cardinalities n_e^{pre} and n_e^{post} after dropping any missing months. We compute the pre- and post-event averages:

$$\bar{y}_e^{\text{pre}} = \frac{1}{n_e^{\text{pre}}} \sum\nolimits_{t \in \mathscr{T}_e^{\text{pre}}} y_t, \; \bar{y}_e^{\text{post}} = \frac{1}{n_e^{\text{post}}} \sum\nolimits_{t \in \mathscr{T}_e^{\text{post}}} y_t, \tag{9}$$

and summarise changes using the absolute difference $\Delta_e = \bar{y}_e^{\text{post}} - \bar{y}_e^{\text{pre}}$ and the percentage change:

$$\%\Delta_e = 100 \times \frac{\bar{y}_e^{\text{post}} - \bar{y}_e^{\text{pre}}}{\bar{y}_e^{\text{pre}}}. \tag{10}$$

To facilitate comparability over time, we also standardise y_t using full-sample moments, $z_t = (y_t - \mu)/\sigma$, where μ is the sample mean and σ the population standard deviation of $\{y_t\}$ over the entire study period; the z-score at the event month is reported as z_{t_e}. Within the combined window $\{t_e - w, \ldots, t_e + w\}$, we record the peak month $\hat{t}_e = \text{argmax}_t y_t$, the associated peak count $y_{\hat{t}_e}$, and the signed offset in months $m_e = (\hat{t}_e - t_e)$, where $m_e < 0$ indicates that coverage peaked before the event. By default, the analysis uses a curated set of widely cited milestones in climate policy and supervision (e.g., Paris Agreement adoption, TCFD recommendations, NGFS-related announcements), but the event list is fully user-configurable via a two-column CSV (event,date). The procedure outputs (i) a compact summary table with one row per event containing $\bar{y}_e^{\text{pre}}$, $\bar{y}_e^{\text{post}}$, Δ_e, $\%\Delta_e$, z_{t_e}, $y_{\hat{t}_e}$, and m_e; and (ii) a single timeline figure plotting monthly counts with vertical markers and labels at each t_e. The objective of event alignment analysis is to quantify the relationship between major climate policy and regulatory milestones and media attention, revealing whether news coverage anticipates or follows key events, which events generate sustained attention shifts versus temporary spikes, and how the banking sector's responsiveness to climate policy has evolved over time. Results from this analysis are presented in 4.10 Event alignment.

3.8. Discourse network

We construct a discourse network by dictionary-matching named actors and themes in each article and aggregating their co-occurrences at the *article* level. Let $\mathscr{B}$ denote bank aliases (e.g., *JPMorgan*, *HSBC*), $\mathscr{R}$ supervisory bodies (e.g., *ECB*, *Federal Reserve*), and $\mathscr{T}$ climate-finance themes (e.g., *net zero*, *stress test*, *disclosure*). For each article a, case-insensitive regular expressions with word boundaries are applied to the full text to obtain the present set $S_a \subseteq \mathscr{B} \cup \mathscr{R} \cup \mathscr{T}$. We then form an undirected co-occurrence graph on the union of nodes with weighted edges

$$w_{ij} = \sum\nolimits_a \mathbf{1}\{i \in S_a, j \in S_a\}, \tag{11}$$

and retain edges with $w_{ij} \geq \tau$ (default $\tau = 2$) to reduce visual clutter. From this full tripartite graph we derive (i) an *actor–actor* projection on $\mathscr{A} = \mathscr{B} \cup \mathscr{R}$ and (ii) an *actor–theme* bipartite graph between $\mathscr{A}$ and $\mathscr{T}$. For each graph $G = (V, E)$ we compute standard weighted metrics: weighted degree $d_i = \sum_j w_{ij}$, betweenness centrality (with edge weights), and eigenvector centrality (when numerically stable). Community structure is estimated via a greedy modularity algorithm on G using the observed weights. For presentation we use a Kamada–Kawai layout; nodes are sized by $\sqrt{d_i}$, and actor types are visually distinguished (banks vs. regulators) (Borgatti and Everett, 1997). The objective of discourse network analysis is to map the relational structure of climate risk discourse by identifying which banks and regulators are most frequently discussed together, revealing the institutional architecture of climate risk governance conversations. This analysis identifies central actors, community structures, and the pathways through which climate risk narratives flow between regulatory and banking domains. Results are presented in 4.11 Discourse Network.

4. Results

4.1. Documents length and distribution

The histogram (Fig. 1) reveals that most articles cluster in the range of 240–1947 words, with a notable peak occurring between 1500 and 1750 words. This suggests that the majority of climate risk coverage in banking takes the form of longer-length articles to provide substantive analysis. Upon further investigation, we noticed that some of the longer articles are the ones which have been also posted in major news outlets such as The Wall Street Journal Online's Real Time Economics for example and the short articles were brief news. This word count distribution reflects how the financial media approaches climate risk coverage - predominantly through standard analyses, and deep dives warranted by the complexity and significance of the topic.

The sample begins in 2009 (Fig. 2) with relatively modest but volatile coverage, showing some early spikes of interest that quickly subsided. These early peaks might represent initial reactions to the growing awareness of climate risks in the aftermath of the global financial crisis, when the banking sector was generally under increased scrutiny. Through the middle period (2013–2017), we see a relatively stable but low baseline of coverage, suggesting that climate risk remained a consistent but not dominant concern in banking discussions. The dramatic peak in 2023 likely corresponds to COP26 and the Glasgow Financial Alliance for Net Zero (GFANZ) commitments, increased regulatory pressure from central banks on climate stress testing, the ECB's climate stress test results and new supervision requirements, the SEC's proposed climate disclosure rules affecting financial institutions. The pattern in recent years (2021–2023) shows not just higher overall coverage but also greater volatility, with frequent peaks and troughs. This suggests that climate risk has become deeply embedded in banking discourse, with coverage driven by a mix of regular reporting cycles and reaction to specific events or developments. Most importantly, the overall trend line, despite its fluctuations, shows a clear upward trajectory over the 14-year period. This evolution reflects the banking industry's growing recognition of climate risk as a fundamental concern rather than a peripheral issue, with coverage becoming both more frequent and more intense in recent years.

4.2. Insights on countries

The United States dramatically dominates the conversation, with nearly 7000 mentions - more than triple its nearest peer, the United Kingdom (2075 mentions) (Fig. 3(a)). This dominance is particularly visible in the time series visualization, where the US line consistently runs above other countries and shows several dramatic spikes, particularly in 2022–2023. This reflects the US's outsized influence in global banking and its increasingly proactive stance on climate risk, especially

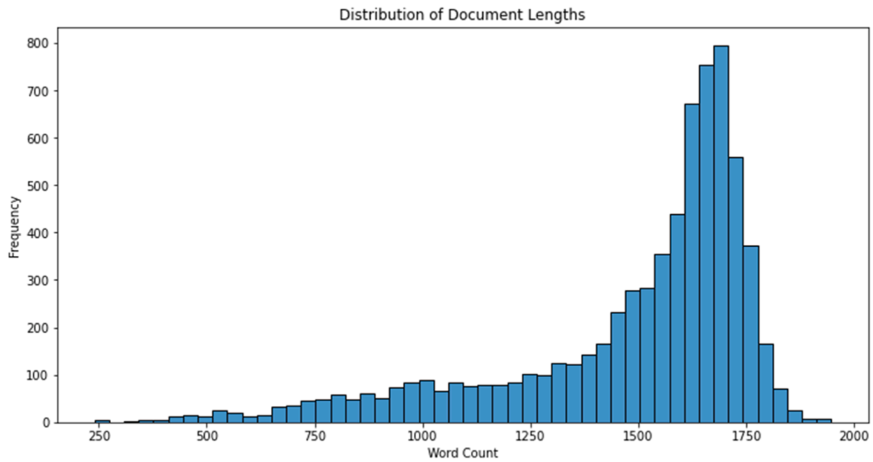


Fig. 1. Distribution of word counts.

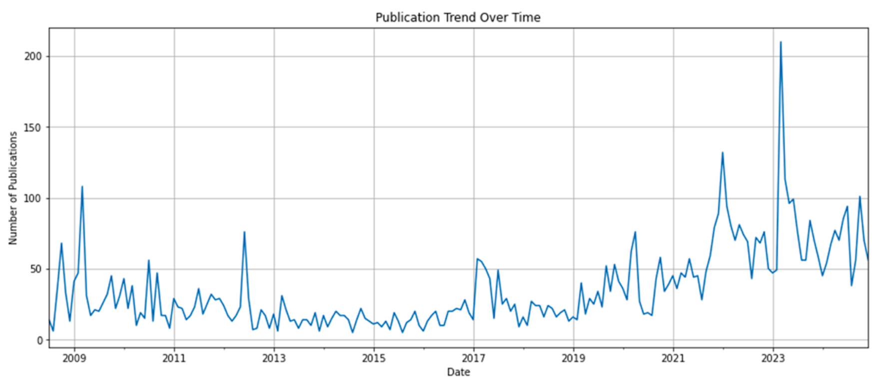


Fig. 2. Number of news articles per year.

following the Biden administration's emphasis on climate-related financial risks and the SEC's proposed climate disclosure rules.

The United Kingdom's strong second-place position (2075 mentions) isn't surprising given London's role as a global financial hub and the Bank of England's early leadership in addressing climate risks. The UK line shows consistent activity throughout the period, with notable increases coinciding with key policy initiatives like the UK's Green Finance Strategy and mandatory TCFD reporting requirements.

China's third-place ranking (1754 mentions) tells an interesting story about the emerging importance of climate risk in Asian banking. The time series shows China becoming more prominent in the discourse around 2020–2021, likely reflecting its announcement of carbon neutrality goals and the expansion of its green finance initiatives.

Australia's significant presence (1079 mentions) likely reflects its unique position as both a major coal exporter and a country highly vulnerable to physical climate risks, with its banking sector increasingly caught between these competing pressures.

The clustering of mentions for Japan, South Africa, Russia, India, Germany, and Canada (ranging from 886 to 538 mentions) suggests a second tier of attention, with each country's coverage often spiking around specific events or policy announcements. For instance, Japan's coverage often aligns with its central bank's climate initiatives, while South Africa's mentions frequently correlate with discussions about transition risks in emerging markets.

The time series reveals an interesting pattern where country mentions have become more volatile since 2020, however there is a clear gap between US and the rest of world in recent years (Fig. 3(b)).

4.3. Most mentioned banks

As can be seen in Fig. 4, UBS stands out dramatically at the top of the chart with a TF-IDF score of 516.93, nearly double that of its nearest peer. This commanding lead suggests that the Swiss banking giant has positioned itself as a leading voice in climate risk discussions, likely through its substantial commitments to sustainable finance and detailed climate risk reporting initiatives.

The Federal Reserve's strong second-place showing (265.64) reflects its dual role as both a regulator and policy leader in climate risk management. Its prominent position underscores the growing importance of central banks in shaping the climate risk agenda for the entire banking sector.

A cluster of major U.S. investment banks follows, with JPMorgan Chase (199.08), Citigroup (173.38), and Goldman Sachs (106.42) all featuring prominently. Their strong presence reflects both their market leadership and their significant efforts to integrate climate risk into their operations and client services.

Deutsche Bank's fifth-place position (136.61) marks it as one of Europe's most prominent voice in the climate risk conversation outside of UBS, likely due to its significant role in sustainable finance and its early adoption of climate risk assessment frameworks.

The presence of multilateral institutions like the European Central Bank (56.67), World Bank (46.15), and IMF (38.90) highlights the crucial role of international financial organizations in developing and implementing climate risk frameworks.

Interestingly, Asian banks appear lower in the rankings, with Bank of China (16.42) and Mitsubishi UFJ (15.84) at the bottom of the top 20, suggesting that while these institutions are engaged in climate risk

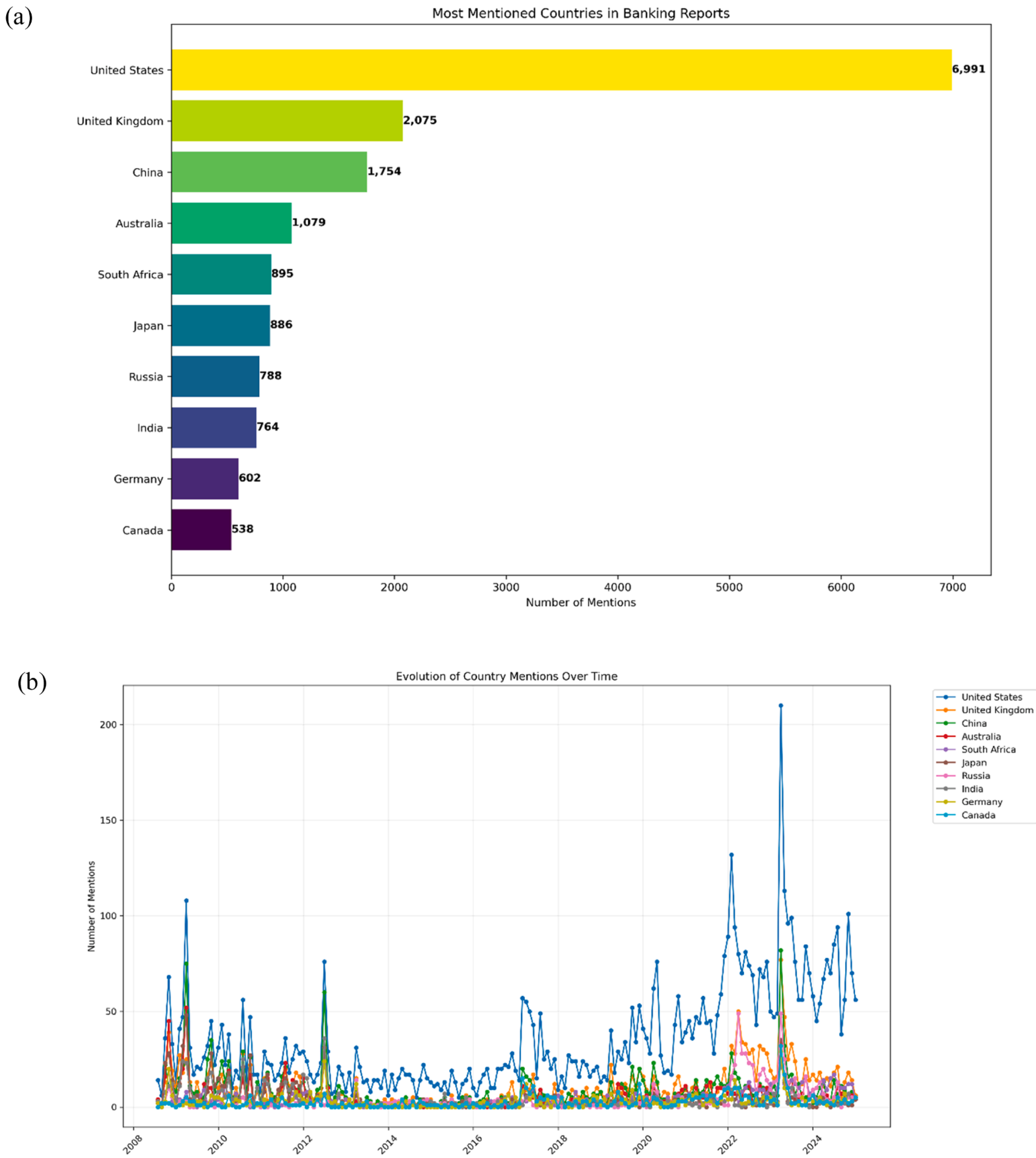


Fig. 3. (a) Most mentioned countries, (b) Evolution of country mentions.

discussions, their voice in global coverage is still developing.

The mix of traditional commercial banks (Bank of America, Wells Fargo), investment banks (Morgan Stanley, Goldman Sachs), and global financial institutions creates a comprehensive picture of how climate risk is being addressed across different types of financial institutions and jurisdictions.

4.4. Salient financial sectors

Risk Management stands clearly at the forefront with the highest TF-IDF score of 148.47, and its dominance is particularly visible in the time series visualization (Fig. 5(a)). This leadership position isn't surprising, as climate risk has fundamentally become a risk management issue for financial institutions. The time series shows dramatic spikes in risk management mentions around 2022–2023, likely corresponding to major regulatory developments and increased scrutiny of banks' climate risk frameworks.

Insurance emerges as the second most prominent sector (126.25), reflecting its position at the frontline of climate change impacts. The sector's coverage has remained relatively consistent over time but shows increasing volatility in recent years, suggesting growing attention to climate-related insurance challenges and opportunities.

Capital Markets (99.12) and Asset Management (85.91) follow closely, with both sectors showing interesting patterns in the time series. Capital Markets exhibited notable early activity around 2009–2012, possibly reflecting early discussions about green bonds and climate-related financial products. The time series shows several significant spikes for Capital Markets, particularly in recent years, suggesting periodic surges of innovation in climate-related financial instruments.

Commercial Banking (74.53) and Investment Banking (70.14) occupy the middle tier, showing a steady presence throughout the period but with less dramatic fluctuations than the top sectors. Their

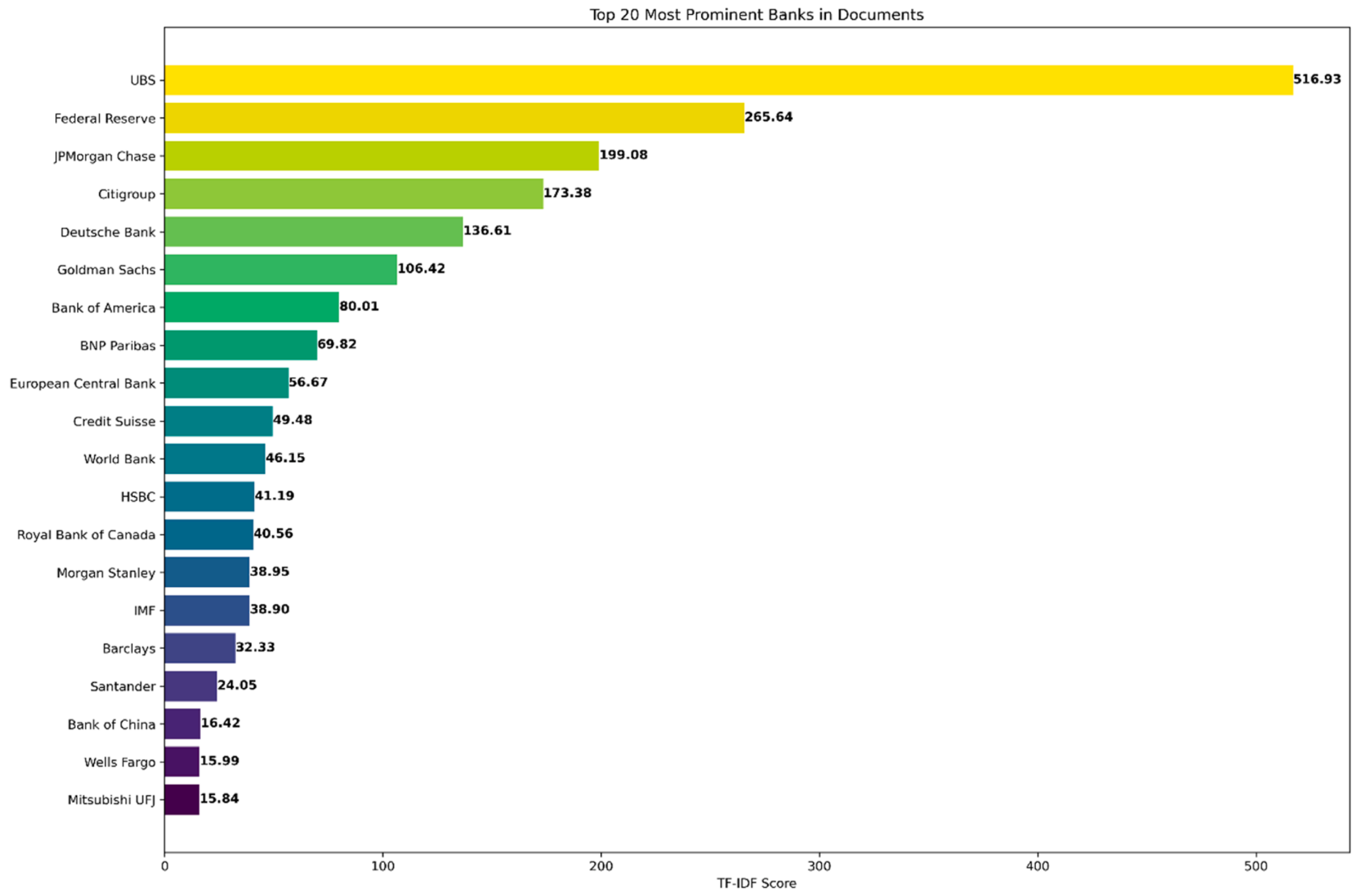


Fig. 4. Most mentioned banks.

relatively stable coverage suggests these traditional banking sectors are consistently engaged with climate risk issues, though perhaps in a more measured way than the more specialized sectors.

The lower tier, including Retail Banking (25.42), Private Equity (19.36), and Fintech (11.89), shows significantly less prominence in climate risk discussions. An interesting pattern emerges in the time series where all sectors show increased activity and volatility post-2020 (Fig. 5(b)), suggesting a broader mainstreaming of climate risk considerations across the financial industry. The dramatic spikes in Risk Management and Capital Markets during this period often coincide, indicating moments of industry-wide attention to major climate risk developments.

4.5. Insights on investment and regulation related discourse

As shown in Fig. 6, investment-related terms (shown in blue) consistently demonstrate higher volatility and more dramatic spikes throughout the period, suggesting that market responses to climate risk tend to generate more intense media attention. The pattern shows several notable peaks, with particularly dramatic spikes around 2009, 2012, and a remarkable surge in 2022, where the TF-IDF score reached above 70. These spikes often coincide with major climate finance initiatives or significant market developments.

In contrast, regulation-related terms (shown in red) display a more measured and gradually increasing presence. The regulatory discourse maintains a lower but more consistent baseline, suggesting a steady evolution of climate risk governance in banking. While the regulatory line shows less dramatic peaks, it demonstrates a gradual upward trend, particularly noticeable from 2016 onwards, reflecting the increasing institutionalization of climate risk considerations in banking supervision.

The persistent gap between investment and regulatory coverage might suggest that market-driven responses to climate risk continue to outpace regulatory frameworks, though the gradually rising regulatory baseline indicates growing supervisory maturity in this area.

4.6. Climate risk related themes

ESG (Environmental, Social, and Governance) emerges as the dominant theme with 7860 mentions, towering over other categories (Fig. 7). This overwhelming prominence suggests that banks are increasingly viewing climate risk through the broader lens of ESG frameworks, integrating it into their overall sustainability and governance strategies.

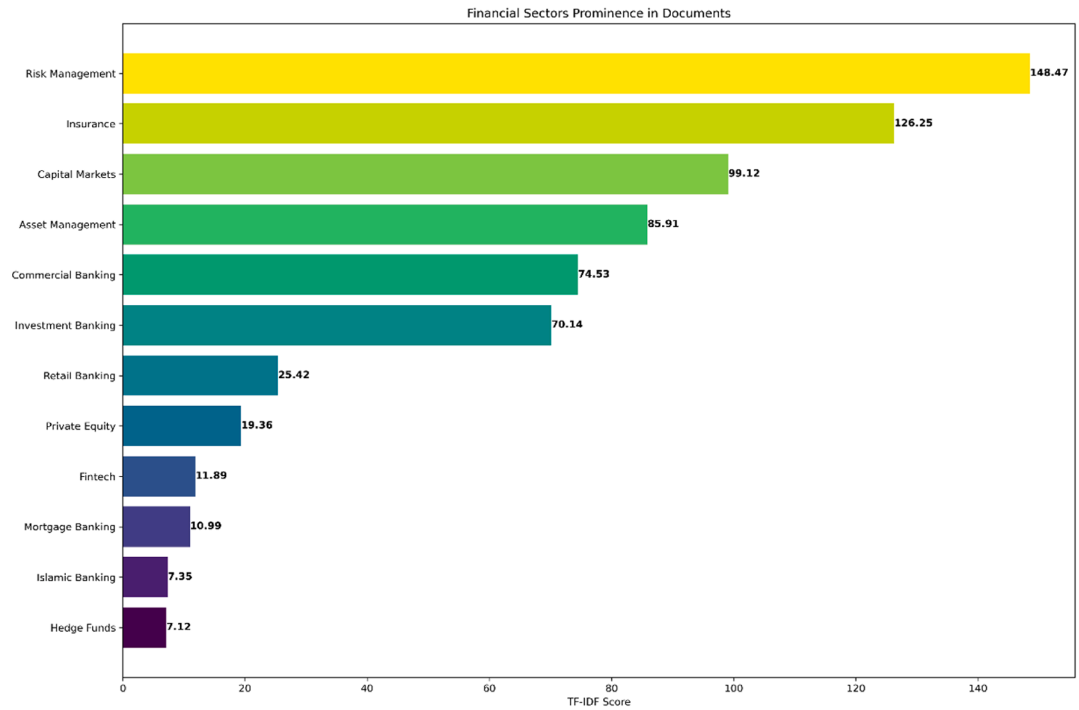
Sustainability follows as the second most frequent theme with 3353 mentions, indicating its establishment as a core concept in banking risk discussions. The significant gap between ESG and sustainability might suggest that while sustainability is important, banks are taking a more comprehensive ESG approach to climate-related challenges.

Carbon appears as the third most prominent theme (1741 mentions), reflecting the banking industry's focus on carbon-related risks and opportunities, likely driven by increasing pressure for carbon accounting and the growth of carbon markets.

Climate change and emissions (1149 and 838 mentions respectively) form a middle tier of frequency, suggesting these more direct environmental terms are often discussed in more specific contexts rather than as overarching frameworks like ESG or sustainability.

Interestingly, more technical risk-related terms like "climate risk," "environmental risk," "transition risk," and "physical risk" show relatively low frequencies (137, 45, 52, and 24 mentions respectively). This might indicate that while these concepts are crucial to risk management, they're often discussed within broader frameworks rather than as

(a)



(b)

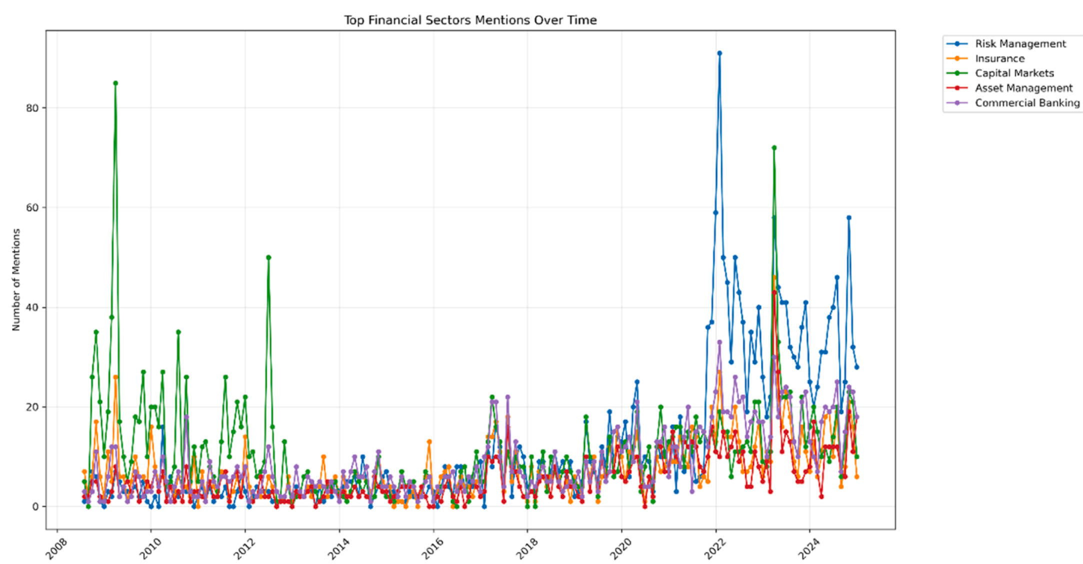


Fig. 5. (a) Sector mentions, (b) Sector evolution over time.

standalone topics.

The emergence of newer themes like "net zero" (201 mentions) and "biodiversity" (65 mentions) with lower but notable frequencies suggests these are emerging areas of concern that are gaining traction in banking risk discussions.

This distribution provides a clear picture of how banks are approaching climate risk - primarily through comprehensive ESG frameworks, while maintaining focus on specific environmental metrics and gradually incorporating emerging environmental concerns.

4.7. Cooccurrence analysis of ESG

Fig. 8 shows that "Bank" and "rating" emerge as the dominant terms throughout the period, but their evolution tells an interesting story. Banking terms show consistently high intensity (red) throughout, starting at 40.8 in 2008 and maintaining strong presence, peaking at 106.7 in 2016. This suggests that banking institutions have been central to the ESG conversation from the beginning.

The "rating" category shows perhaps the most dramatic transformation, starting modestly at 4.6 in 2008 but intensifying significantly over time, reaching its peak of 211.4 in 2022 (shown in deep burgundy). This dramatic increase reflects the growing importance of rating considerations in ESG frameworks and the rise of ESG-focused investment strategies. This is supported by similar trend for “fitch” as one of the most prominent credit rating agencies, starting at nearly zero in 2008 but gradually gaining prominence, particularly post-2019. Another closely related theme, "credit" show a steady increase in prominence over time, becoming particularly significant after 2020 (reaching 41.5 in 2022), indicating the growing integration of ESG considerations into credit risk assessment and lending decisions.

The "risk" category shows notable fluctuations but maintains consistent presence throughout the period, with particular intensity in recent years (peaking at 29.7 in 2022), reflecting the mainstreaming of ESG risk considerations in banking operations.

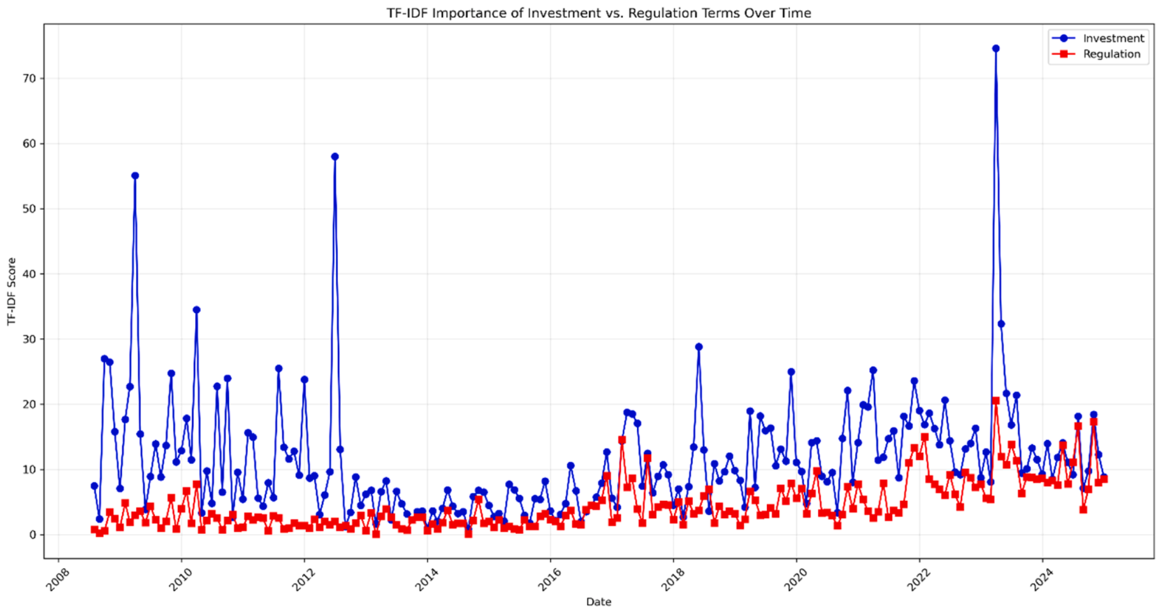


Fig. 6. Relative importance of investment vs regulation based on TF-IDF scores.

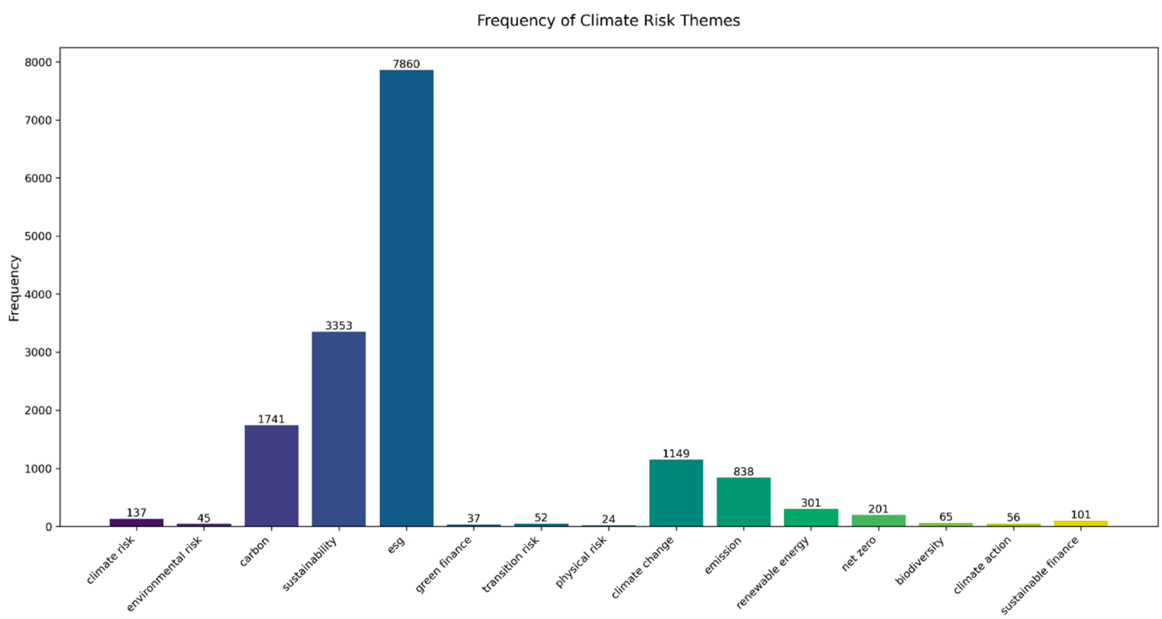


Fig. 7. Frequency of climate risk themes.

An interesting pattern emerges in the "business" and "financial" categories, which show relatively stable moderate intensity throughout the period, suggesting these represent foundational aspects of ESG implementation in banking.

The most recent years (2020–2024) show the highest overall intensity across multiple categories, indicating a maturation and broader integration of ESG considerations across various aspects of banking operations.

4.8. Sentiment analysis

The overall sentiment distribution (Fig. 9) reveals a decidedly positive skew, with positive sentiment dominating at around 4000 mentions, significantly outweighing both negative (approximately 1500) and neutral (about 1000) sentiments. This suggests that the banking industry's discourse around climate risk tends to focus on opportunities, solutions, and proactive measures rather than threats and challenges.

The continuous sentiment score distribution shows high confidence in the sentiment classifications with most scores clustering between 0.9 and 1.0. The positive sentiment curve (blue) shows the highest peak, confirming the strong positive bias in the coverage, while neutral and negative sentiments show lower but still significant densities.

Looking at the temporal evolution of sentiments from 2009 to 2024 (Fig. 10), we can observe that in early period (2009–2015) climate risk had a more modest coverage overall with occasional positive spikes with relatively balanced distribution between sentiments. For middle Period (2016–2020) sentiment show a modest distribution with the least gap between positive, negative and neutral. And finally for recent Period (2021–2024) we observe the most dramatic increase in coverage volume and sentiment volatility with notable positive sentiment spikes, particularly in 2023 (reaching above 120 counts). More pronounced separation between positive and negative/neutral coverage can also be seen in

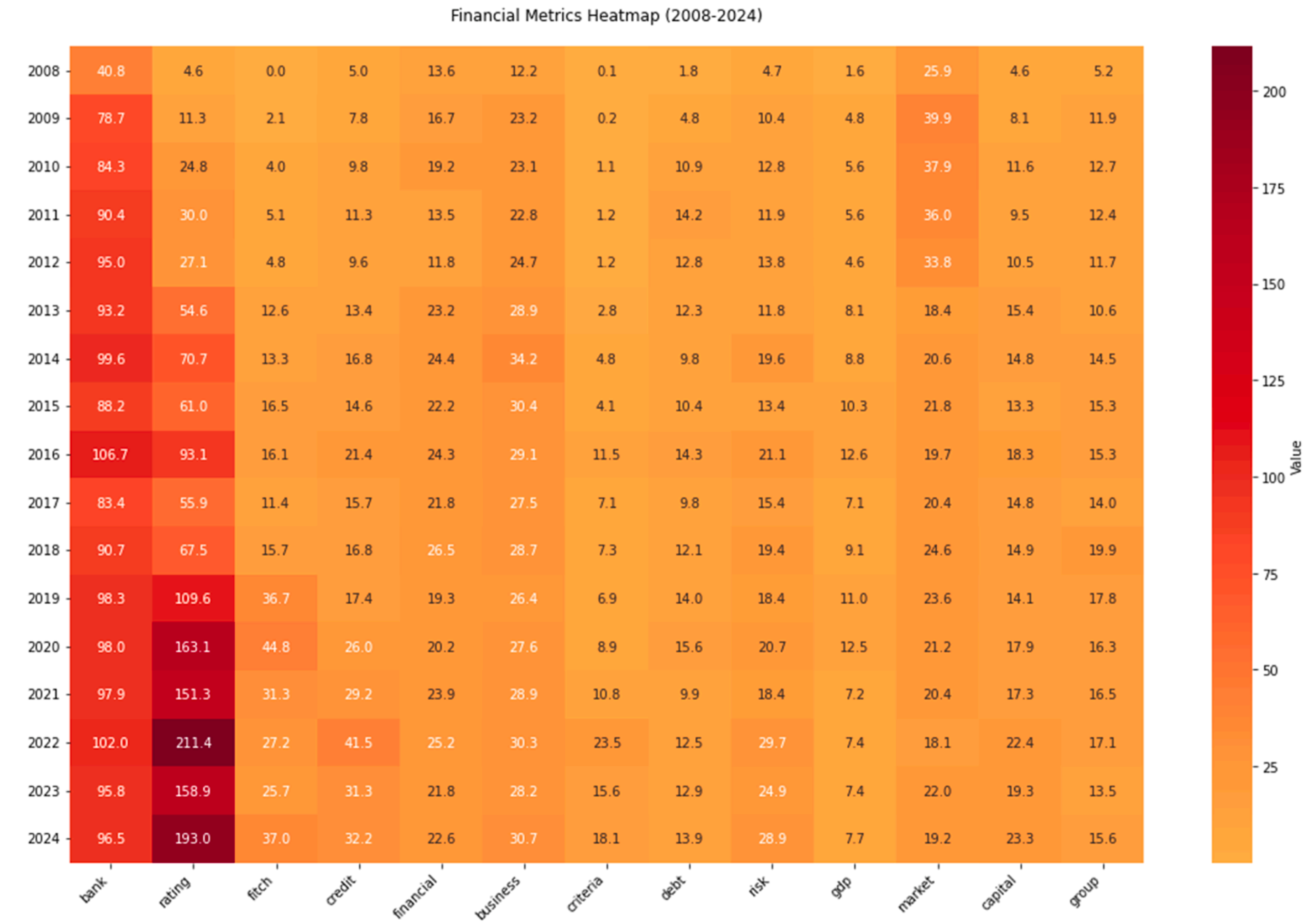


Fig. 8. Terms associated with ESG.

this period.

The most striking feature is the dramatic spike in positive sentiment during 2023, possibly reflecting optimism about new climate-related banking initiatives or successful adaptation strategies. Throughout the entire period, positive sentiment maintains leadership, though with increasing volatility in recent years, suggesting growing complexity in how the banking sector engages with climate risk issues.

4.9. Topic modeling

Through UMAP dimensionality reduction, eight distinct clusters emerge, each representing a significant aspect of the climate risk discourse in banking (Fig. 11 and Table 1).

The analysis identifies Corporate Financial Operations as a central node in the thematic network, exhibiting high density and centrality in the visualization. This positioning suggests its fundamental role in operational implementation of climate risk strategies. The cluster's strong interconnections with adjacent themes indicate its role as a primary driver of climate risk integration in banking practices.

Sovereign Debt & Governance emerges as a distinct thematic domain, spatially separated from operational clusters. This separation is theoretically significant, suggesting a clear delineation between policy-level discussions and operational implementation. The cluster's positioning reflects the regulatory and governance frameworks that shape climate risk management in banking.

Credit Rating & ESG Assessment demonstrates significant intermediary characteristics, bridging the gap between theoretical frameworks and practical implementation. This cluster's position suggests its crucial role in translating climate risk considerations into quantifiable metrics for financial decision-making.

Banking Performance & Disclosure exhibits strong cohesion as a cluster while maintaining connections to both operational and market-oriented themes. This positioning reflects the growing emphasis on transparency and standardization in climate risk reporting within the banking sector.

The Market News & Analysis cluster displays broader distribution across the thematic space, indicating its pervasive influence in contextualizing climate risk across multiple domains. This diffusion suggests its role in interpreting and disseminating climate risk information across the banking sector.

Rating Methodology & Criteria forms a discrete cluster with well-defined boundaries, indicating the technical specificity of climate risk assessment frameworks. Its positioning suggests a specialized domain that influences both assessment protocols and performance metrics.

Macroeconomic & Fiscal Indicators demonstrates significant connectivity across clusters while maintaining its distinct identity. This positioning reflects its role in providing broader economic context to climate risk considerations in banking.

Investment Products & ETFs appears as a peripheral yet well-defined cluster, suggesting the emergence of specialized financial instruments addressing climate risk. Its position indicates the practical manifestation of climate risk strategies in marketable financial products.

The three-dimensional visualization further elucidates these relationships, revealing hierarchical structures and interconnections not immediately apparent in two-dimensional analysis. This additional dimension provides insights into the complexity of thematic

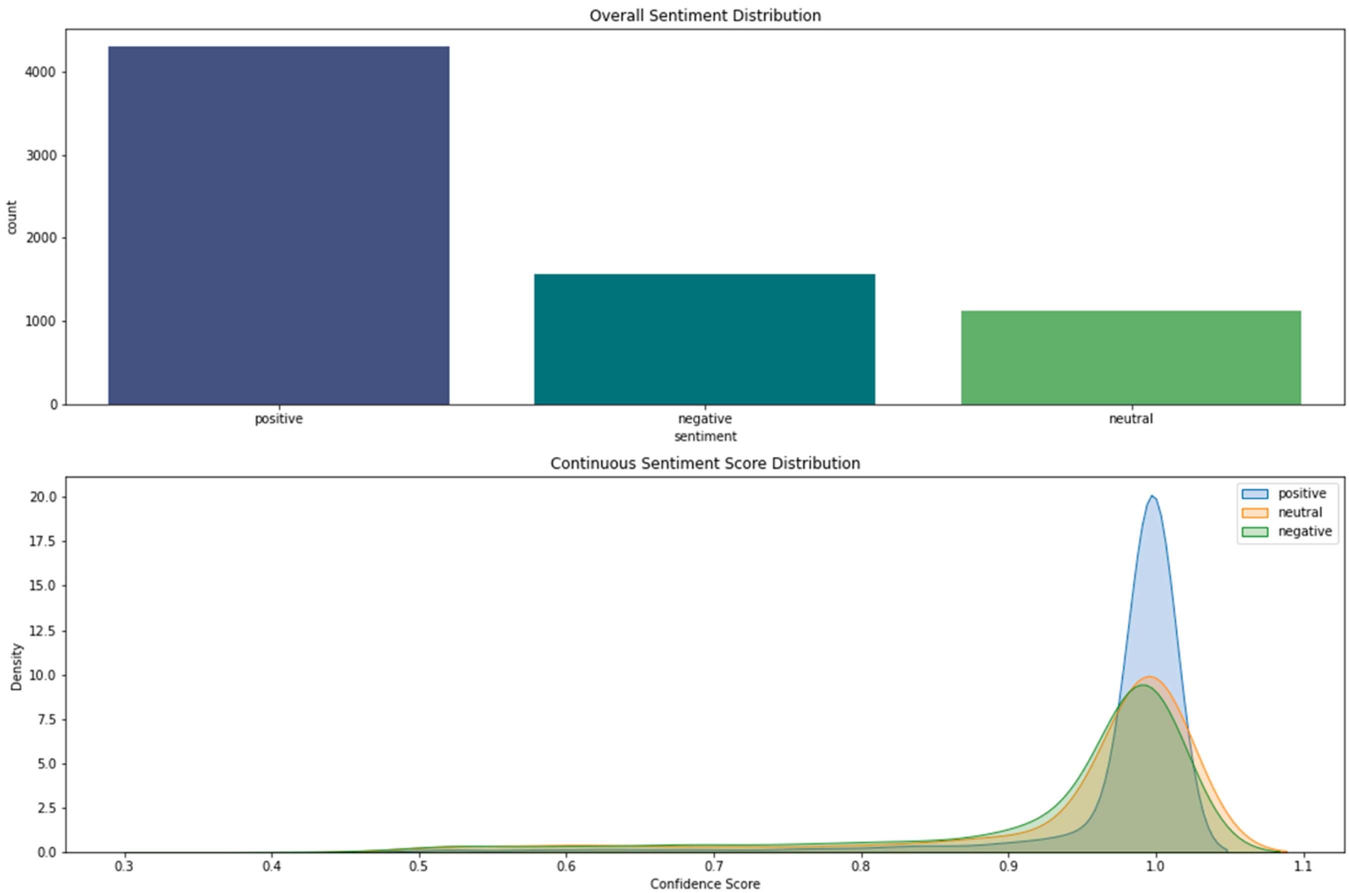


Fig. 9. Overall sentiment distribution and continuous sentiment distribution.

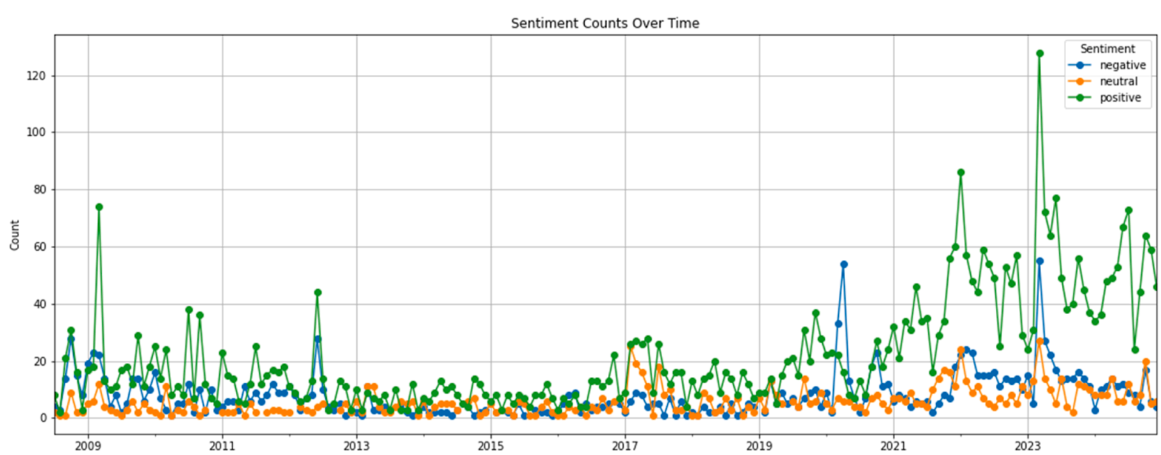


Fig. 10. Sentiment counts over time.

relationships within climate risk discourse in banking.

4.10. Event alignment

Overall, banking-climate coverage clearly clusters around major global and supervisory milestones, but the timing is asymmetric; across 25 anchors the median peak occurs around 2 months before the event (months-to-peak median $= -2$), indicating anticipatory reporting. Even so, a few headline events drove powerful post-event surges. COP26 (Oct-2021) doubles average monthly coverage within its ±6-month window (+100%, 44→89), peaking + 3 months. The Basel Committee climate-risk consultation (Nov-2021) is nearly identical (+93.6%, 47→91; peak +2 months). The Bank of England CBES results produce the largest single peak in the sample (210 articles; peak +2 months) and a pronounced level shift (+62.4%, 63→102) (Table 2).

By contrast, several 2022 rule-making disclosures show sharp event-month spikes but lower post-event averages. The SEC climate disclosure proposal (Mar-2022) records the highest event-month z-score (1.59), yet its post window averages −16.8% below the pre window (84→70) with the peak 2 months before the announcement. Similar patterns hold for the ECB climate stress-test results (Apr-2022) and the Basel climate-risk principles (Jul-2022) (event-month $z \approx 1.2$–1.3; peaks 3–6 months

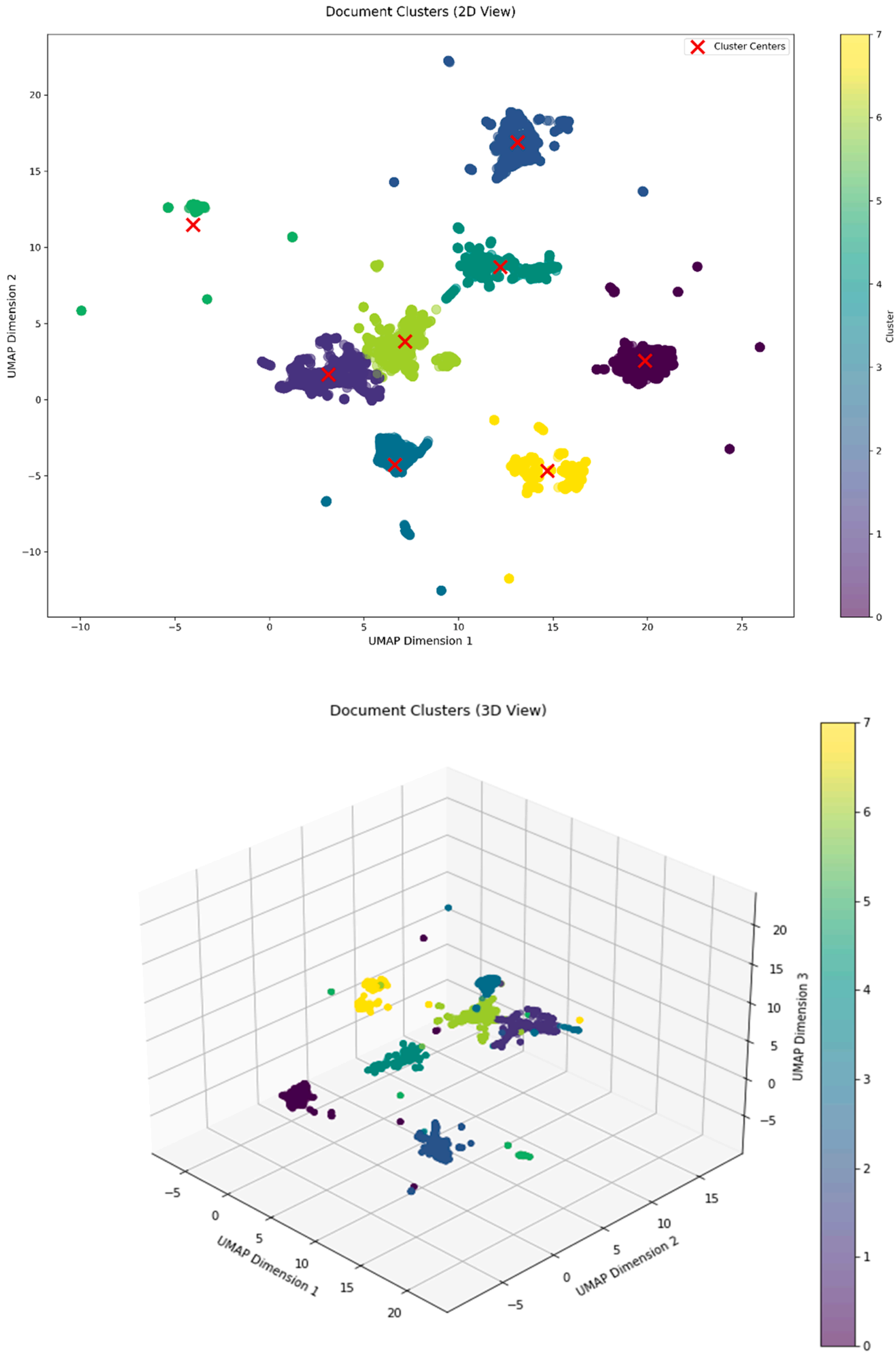


Fig. 11. 2D and 3D visualization of the clusters TF-IDF.

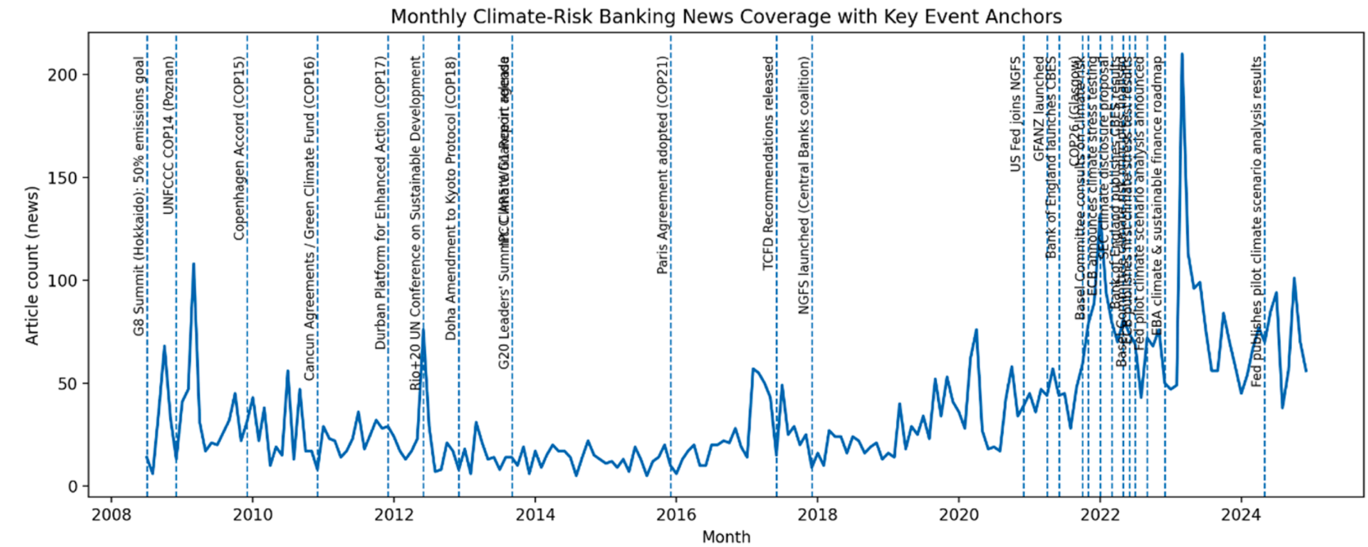


Fig. 12. Monthly news coverage with key event anchors.

Table 1

Major topics (clusters) in the news about climate risk in banking industry.

Cluster	Title	Prominent Keywords	Description
Cluster 0	Corporate Financial Operations	financial, performance, revenue, operations, investment, portfolio, assets, management, strategy, business	Focuses on how climate risk affects day-to-day banking operations, including financial performance metrics, investment strategies, and asset management decisions. Covers operational adaptation to climate considerations and integration into core business processes.
Cluster 1	Sovereign Debt & Governance	government, policy, sovereign, debt, public, national, regulatory, framework, governance, institutional	Addresses government-level climate risk policies, sovereign debt implications, and national regulatory frameworks. Includes discussions of central bank policies, government green bonds, and institutional governance structures for climate risk management.
Cluster 2	Credit Rating & ESG Assessment	rating, credit, ESG, assessment, evaluation, scoring, methodology, criteria, standards, analysis	Centers on the evolution of credit rating methodologies to incorporate climate risks and ESG factors. Covers rating agency approaches, scoring mechanisms, and standardization efforts in climate risk assessment for financial institutions.
Cluster 3	Banking Performance & Disclosure	disclosure, reporting, transparency, performance, metrics, indicators, compliance, standards, communication	Focuses on climate risk disclosure requirements, reporting standards, and performance measurement frameworks. Includes regulatory compliance, and standardization of climate risk communication in banking.
Cluster 4	Market News & Analysis	market, news, analysis, trends, developments, industry, sector, commentary, insights, outlook	Encompasses general market commentary, industry analysis, and news reporting on climate risk developments in banking. Provides contextual coverage and interpretive analysis of market trends and regulatory developments.
Cluster 5	Rating Methodology & Criteria	methodology, criteria, approach, framework, standards, guidelines, procedures, technical, specification, requirements	Addresses the technical aspects of climate risk assessment methodologies, including detailed frameworks, evaluation criteria, and standardized approaches used by financial institutions and rating agencies.
Cluster 6	Macroeconomic & Fiscal Indicators	economic, fiscal, macroeconomic, growth, inflation, monetary, policy, central, bank	Covers broader economic implications of climate risk, including macroeconomic indicators, monetary policy considerations, and fiscal policy responses. Addresses systemic economic risks and central banking approaches to climate challenges.
Cluster 7	Investment Products & ETFs	investment, products, ETF, funds, green, sustainable, portfolio, allocation, returns, instruments	Focuses on climate-aligned investment products, including green bonds, sustainable ETFs, and specialized financial instruments. Covers product development, market performance, and investment allocation strategies in climate-related financial products.

before; post windows −23.3% and −28.8%, respectively). Earlier UNFCCC milestones (2008–2013) show more modest, short-lived upticks (e.g., COP14 +36.4%) relative to the larger, more sustained surges centered on 2021–2023. In total, 12 of 26 anchors raise the post-event average while 13 show neutral/negative drift.

4.11. Discourse network

The discourse network is anchored by two regulatory hubs (the U.S. Federal Reserve and the European Central Bank) whose weighted-degree centralities (3200 and 2143, respectively) (Table 3) exceed most of those of all other actors. The Fed–ECB tie is also the single strongest edge (weight = 297), indicating their role as the principal coordinators of the global conversation on climate-related financial risk.

Surrounding this regulatory core is a group of systemically important banks, mainly U.S. and European, that function as the primary conduits into supervisory discourse (Fig. 13). Citigroup is the most frequently discussed bank (weighted degree = 2592) and maintains strong ties to both the Federal Reserve (276) and the ECB (143). It is closely followed by JPMorgan (1998), UBS (1825), Morgan Stanley (1764), and Goldman Sachs (1622). Their high eigenvector centralities confirm that these institutions sit at the heart of the narrative, repeatedly paired with regulators as both subjects of oversight and partners in implementation.

Community structure reveals a distinct European bloc comprising the ECB, the Bank of England, and leading European banks (HSBC (1719), Barclays (1719), BNP Paribas (938), and Société Générale (811)). Dense intra-European ties (e.g., Barclays–HSBC, 145; ECB–Bank of England, 143) (Table 4) indicate a regionally focused discourse

Table 2
Event alignment top 10 results summary.

Event	event_date	pre_avg_count	post_avg_count	Change (%)	peak_month	zscore
COP26 (Glasgow)	10/31/2021	44.33	88.83	100.4	1/1/2022	0.84
Basel Committee consults on climate risk	11/16/2021	46.83	90.67	93.6	1/1/2022	1.56
ECB announces climate stress testing	1/27/2022	58	88.5	52.6	1/1/2022	3.45
US Fed joins NGFS	12/15/2020	31.5	44.67	41.8	10/1/2020	0.13
EBA climate & sustainable finance roadmap	12/13/2022	67	94.17	40.5	3/1/2023	0.52
UNFCCC COP14 (Poznan)	12/1/2008	31.4	42.83	36.4	3/1/2009	-0.8
Rio+ 20 UN Conference on Sustainable Development	6/20/2012	20.5	26.33	28.5	6/1/2012	1.45
Fed publishes pilot climate scenario analysis results	5/9/2024	61.83	74	19.7	10/1/2024	1.24
Bank of England launches CBES	6/8/2021	44.67	50.5	13.1	11/1/2021	0.31
GFANZ launched	4/21/2021	43.17	44.33	2.7	10/1/2020	0.31

Table 3
Actors by network centrality measures.

node	type	degree_w	eigenvector
citigroup	bank	2592	0.353967
jpmorgan	bank	1998	0.288009
ubs	bank	1825	0.260106
morgan_stanley	bank	1764	0.252469
barclays	bank	1719	0.237388
goldman_sachs	bank	1622	0.235794
hsbc	bank	1531	0.206927
bank_of_america	bank	1523	0.230904
deutsche_bank	bank	1513	0.214696
credit_suisse	bank	1401	0.199841
bnp_paribas	bank	938	0.128876
societe_generale	bank	811	0.109213
wells_fargo	bank	800	0.119428
standard_chartered	bank	723	0.099836
bank_of_china	bank	690	0.092978
mufg	bank	680	0.095886
federal_reserve	regulator	3200	0.40379
ecb	regulator	2143	0.275439
bank_of_england	regulator	1208	0.172906
imf	regulator	912	0.106893
sec	regulator	682	0.110606
world_bank	regulator	441	0.049005
eba	regulator	196	0.023886
basel_committee	regulator	106	0.014044

shaped by EU and UK initiatives.

Overall results exhibit a clear core–periphery pattern. While the Fed, ECB, and major universal banks populate the core, several institutions operate at the margins. The IMF and World Bank form a tight pair (164), frequently referenced together yet less embedded in the day-to-day supervisory conversation; their elevated betweenness suggests a bridging function, linking discussions of climate finance to the core regulatory thread. Standard-setters such as the EBA and the Basel Committee appear less often (low degree) but register high betweenness, acting as strategic connectors when rulemaking or methodological guidance draws multiple larger actors into the same narrative. Finally, Asian banks, including MUFG (680) and Bank of China (690), participate through a more self-contained community, indicating engagement that is visible in the corpus but less interwoven with the Western-centric core.

5. Discussion and conclusions

Our analysis reveals several significant patterns in how the banking sector's approach to climate risk has evolved over the past decade and a half. The temporal analysis demonstrates a clear transformation from sporadic coverage in the early 2010s to intensive engagement by 2023, with particularly notable acceleration during 2021–2023. Event-alignment results add two refinements: first, peaks in attention often precede major regulatory and policy anchors (anticipatory coverage), and second, some anchors generate level shifts rather than short-lived spikes. Notably, the COP26 window is associated with a step-change in baseline attention, whereas supervisory milestones such as the Bank of England's scenario exercise produced sharp but more localized surges. This evolution reflects not just increased attention to climate risk, but a fundamental shift in how banks conceptualize and respond to climate-related challenges. The dramatic surge in ESG-related discussions (7860 mentions) compared to traditional risk categorizations suggests that banks are increasingly approaching climate risk through comprehensive sustainability frameworks rather than treating it as an isolated risk factor.

The geographical analysis reveals significant disparities in climate-risk engagement across regions. The United States' dominance in the discourse (approximately 7000 mentions) compared to other regions suggests that climate-risk integration in banking remains heavily influenced by developed-market practices. The emergence of China as the third most frequently mentioned country (1754 mentions) indicates growing attention in major emerging markets. The discourse network analysis helps explain these patterns: the network is organized around two regulatory hubs (the Federal Reserve and the European Central Bank). This structure suggests that agenda-setting power is concentrated in a small set of authorities and globally active banks, which may accelerate convergence among core jurisdictions while leaving other regions more weakly connected to the evolving policy and practice discourse.

Our sectoral analysis identifies varying levels of engagement across different parts of the banking industry. Risk Management's leading position (TF-IDF score 148.47) indicates that climate considerations have been primarily integrated through risk-management frameworks. By contrast, relatively low engagement from Retail Banking (25.42) and Fintech (11.89) suggests that integration remains uneven. The discourse network provides a plausible mechanism: systemically important banks (e.g., Citigroup, JPMorgan, UBS) occupy brokerage positions between regulators and multiple thematic clusters, helping translate supervisory narratives into risk and finance functions, while more customer-facing or digitally native segments appear less embedded in these translation pathways. This uneven connectivity likely slows diffusion of practice beyond risk and treasury into product design, pricing, and frontline channels.

The sentiment analysis reveals a predominantly positive tendency in climate-risk coverage, with positive mentions (approximately 4000) significantly outnumbering negative ones (approximately 1500). In our topic-modeling analysis, the emergence of Corporate Financial Operations as a central node, bridging various thematic areas, suggests that climate risk has moved beyond purely environmental considerations to become integrated into core banking operations. The clear separation between operational and policy-level clusters (such as Sovereign Debt & Governance) indicates that while strategic frameworks are well developed, implementation challenges persist. Policymakers can increase durability by timing communications around sustained windows (rather than single-day drops) and by partnering with network brokers to diffuse guidance. For banks, strengthening cross-functional links (especially into retail and digital channels) and engaging proactively ahead of known policy anchors may improve preparedness and reduce the

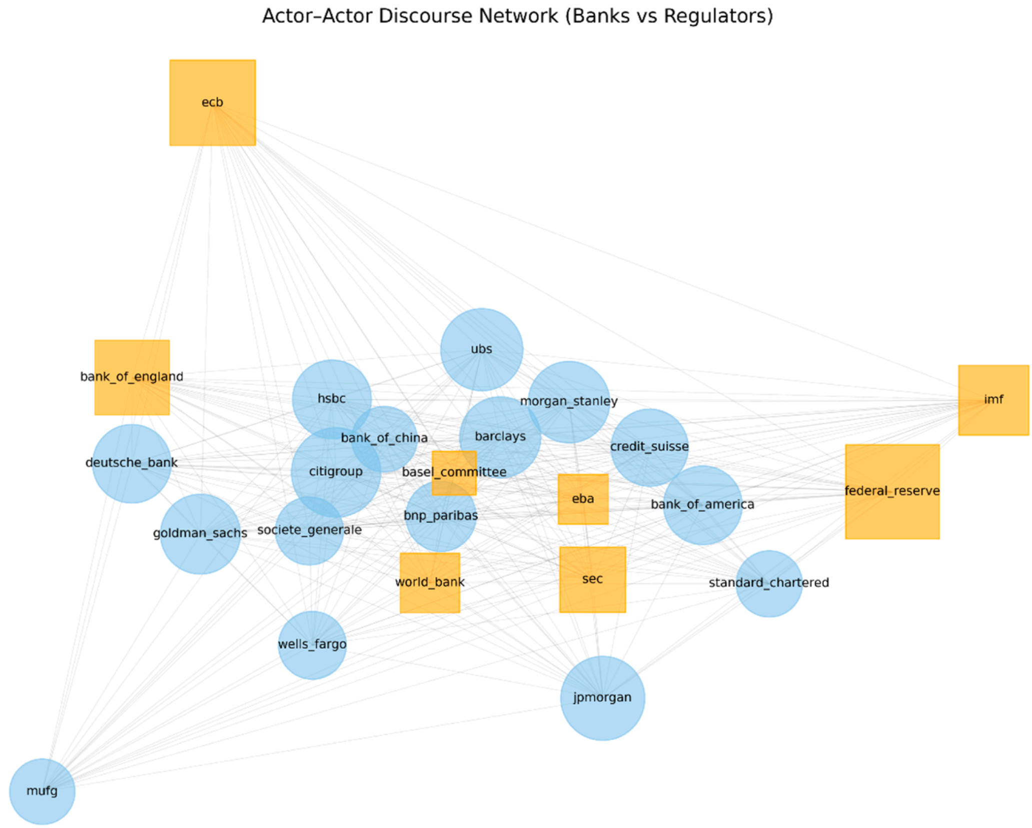


Fig. 13. Banks vs. regulators network.

Table 4
Top 15 actor-actor connection summary.

source	target	weight
ecb	federal_reserve	297
citigroup	federal_reserve	276
federal_reserve	jpmorgan	243
citigroup	morgan_stanley	235
citigroup	jpmorgan	229
bank_of_america	federal_reserve	212
federal_reserve	sec	212
citigroup	ubs	203
bank_of_america	citigroup	201
federal_reserve	ubs	198
federal_reserve	goldman_sachs	192
federal_reserve	morgan_stanley	170
barclays	federal_reserve	166
bank_of_england	federal_reserve	166
imf	world_bank	164

amplitude of sentiment swings.

These findings carry important theoretical implications that extend existing frameworks of climate risk in finance. First, the evolution from standalone climate risk treatment to ESG-embedded approaches directly supports Monasterolo's (2020) argument for integrated frameworks and challenges the adequacy of traditional risk management models identified by Battiston et al. (2021). Our evidence of sector-wide ESG dominance (7860 mentions) empirically validates the theoretical shift from isolated risk factors toward systemic sustainability frameworks that Chenet (2021) conceptualized when distinguishing physical from transition risks. Second, our discourse network findings advance the theoretical understanding of how climate policy shocks propagate through financial networks, as posited by Roncoroni et al. (2021). The hub-and-spoke structure we identify—with the Federal Reserve and ECB as central coordinators and major banks serving as conduits—reveals that climate risk adoption follows network diffusion principles rather than uniform regulatory compliance, extending institutional theory to explain why implementation varies so dramatically across the sector (ECB, 2024; Federal Reserve, 2024). Third, the anticipatory pattern in our event alignment analysis (median peak two months before policy anchors) challenges the reactive assumptions underlying traditional risk management frameworks (Feridun and Güngör, 2020), suggesting that theoretical models must account for the banking sector's forward-looking orientation. The differential persistence of attention—sustained engagement following transformational events like COP26 versus temporary spikes after technical rule-making—indicates that frameworks must distinguish between paradigm-shifting and incremental regulatory change. Finally, our geographical and sectoral variations empirically extend Nieto's (2019) work on differential climate risk integration, demonstrating through discourse analysis that existing theoretical models require modification to account for regional institutional capacity and sector-specific constraints.

For policy and practice, our findings offer several actionable insights. Regulators should recognize that the anticipatory nature of discourse—with attention peaking approximately two months before major announcements—means that communication strategies must prepare for heightened scrutiny before formal policy releases. The differential impact patterns suggest that paradigm-shifting initiatives (such as COP26) warrant extensive stakeholder engagement given their sustained attention effects, whereas technical rule-making may benefit from more targeted, streamlined communication. The significant regional disparities we document, with U.S.-led coverage substantially outpacing other jurisdictions, highlight the need for more coordinated international supervisory frameworks while remaining sensitive to local market conditions. For banking institutions, our discourse network analysis reveals that climate risk leadership requires active participation in regulatory dialogue rather than passive compliance; banks like

Citigroup, JPMorgan, and UBS have positioned themselves as central nodes by serving as bridges between supervisory and market perspectives. The dominance of ESG framing (7860 mentions) indicates that banks should embed climate risk within comprehensive sustainability frameworks rather than treating it as an isolated risk category. Risk management professionals should note that the emergence of eight distinct thematic clusters implies the need for multidimensional assessment frameworks spanning corporate operations to macroeconomic indicators, and the anticipatory discourse patterns suggest implementing forward-looking monitoring systems that track regulatory signals rather than relying solely on reactive approaches. The uneven sectoral engagement—with Risk Management leading (TF-IDF score 148.47) while Retail Banking (25.42) and Fintech (11.89) lag—indicates that diffusing climate risk practices beyond treasury and risk functions into product design, pricing, and customer-facing channels remains a significant implementation challenge.

Several limitations warrant acknowledgment. Most importantly, our reliance on English-language news coverage likely underrepresents perspectives from non-English speaking regions, particularly relevant given significant climate risk initiatives in Europe and Asia where important discussions occur in local languages. Future research should extend this analysis to multilingual sources. Additionally, while our ProQuest corpus aggregates content from diverse originators (see Footnote 4), all articles are channeled through a single publisher (Dow Jones & Company Inc.), which may introduce selection biases that a multi-platform sampling strategy could mitigate. While our NLP methodology captures broad patterns effectively, the use of TF-IDF scores and sentiment analysis may not fully capture the nuanced ways banks discuss and implement climate risk measures. As Cardenas (2024) notes, current tools for climate risk assessment remain inadequate, and this limitation extends to discourse analysis. Future studies could employ more sophisticated natural language processing techniques to capture subtler patterns, complement news analysis with regulatory filings and sustainability reports, and examine how thematic patterns in discourse translate into actual operational changes within banks, following Taylor's (2023) work on implementation challenges.

CRediT authorship contribution statement

Babak Naysary: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ali Edisen:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation. **Amine Tarazi:** Writing – review & editing, Writing – original draft, Validation, Supervision, Investigation, Formal analysis, Conceptualization.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

Not applicable.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

Not applicable.

Appendix A. : Sample News Article

Title: Press Release: Scotiabank top scoring bank in North America for second year in a row, selected to the Dow Jones Sustainability Index
Abstract: None available.
Full text: Scotiabank announced today it secured the top S&P Global ESG Score amongst banks in North America on S&P's 2024 Corporate Sustainability Assessment (CSA). The Bank also maintained its seven-year track record of inclusion in the Dow Jones Sustainability Index (DJSI) North America. Inclusion in the DJSI reflects a global benchmark assessing sustainable business practices that are deemed relevant to generating long-term shareholder value and integrating sustainability considerations into portfolios. The Bank scored 73/100, placing in the top six percent of banks worldwide […]. Scotiabank earned top scores in the categories of Transparency and Reporting and Corporate Governance. The Bank also ranked highly across multiple dimensions, such as Climate Strategy, Business Ethics, Human Capital Management, Talent Planning and Analytics […].
Subject: Banks; Banking industry; Climate change; Sustainability
Business indexing term: Subject: Banks Banking industry; Corporation: Bank of Nova Scotia
Location: North America
Company / organization: Name: Bank of Nova Scotia; NAICS: 522110; Name: New York Stock Exchange–NYSE; NAICS: 523210
Title: Press Release: Scotiabank top scoring bank in North America for second year in a row, selected to the Dow Jones Sustainability Index
Publication title: Dow Jones Institutional News; New York
Publication year: 2024
Publication date: Dec 23, 2024
Publisher: Dow Jones & Company Inc.
Place of publication: New York
Country of publication: United States, New York
Publication subject: Business And Economics
Source type: Wire Feed
Language of publication: English
Document type: News
ProQuest document ID: 3148787705
Document URL: https://www.proquest.com/wire-feeds/press-release-scotiabank-top-scoring-bank-north/docview/3148787705/se-2
Last updated: 2024–12–24
Database: ProQuest One Business

Appendix B. : Algorithm for Extracting Relevant Attributes and Saving to CSV File

Step	Description
1.	Open input text files (14 batch files exported from ProQuest).
2.	For each file, split content using "____" delimiter to separate individual articles.
3.	For each article, extract key attributes: Document ID, Title, Abstract, Full text, Subject, Publisher, Publication date, Publication year, and Document type.
4.	Apply quality control filters: - Remove articles with missing ID or Title - Remove banned document types (advertisements, corrections, notices, etc.) - Remove articles with fewer than 80 words - Remove off-topic articles (subject not containing "bank")
5.	Apply three-stage deduplication: - Stage 1: Remove duplicates by Document ID - Stage 2: Remove duplicates by content hash - Stage 3: Remove duplicates by Title + Publication date + Publisher
6.	Normalize dates to ISO format (YYYY-MM-DD).
7.	Save cleaned dataset to CSV file (processed_documents.csv).

Data Availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Achitouv, I., Gorduza, D., & Jacquier, A. ,2023. Natural language processing for financial regulation. arXiv preprint arXiv:2311.08533.

Altavilla, C., Boucinha, M., Pagano, M., Polo, A., 2024. Climate risk, bank lending and monetary policy. ECB working paper no. 2024/2969.

Alvim, L., Vilela, P., Motta, E., & Milidiú, R.L. ,2010. Sentiment of financial news: a natural language processing approach. 1st Workshop on Natural Language Processing Tools Applied to Discourse Analysis in Psychology, Buenos Aires,

Battiston, S., Dafermos, Y., Monasterolo, I., 2021. In. In: Climate risks and financial stability, 54. Elsevier, 100867.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W., Ng, L.G., Ginhoux, F., Newell, E.W., 2019. Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. 37 (1), 38–44.

Blei, D.M., Kucukelbir, A., McAuliffe, J.D., 2017. Variational inference: a review for statisticians. J. Am. Stat. Assoc. 112 (518), 859–877.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3 (Jan), 993–1022.

Borgatti, S.P., Everett, M.G., 1997. Network analysis of 2-mode data. Soc. Netw. 19 (3), 243–269.

Campiglio, E., Dafermos, Y., Monnin, P., Ryan-Collins, J., Schotten, G., Tanaka, M., 2018. Climate change challenges for central banks and financial regulators. Nat. Clim. Change 8 (6), 462–468.

Cardenas, V. ,2024. Managing financial climate risk in banking services: A review of current practices and the challenges ahead. *arXiv preprint arXiv:2405.17682*.

Chenet, H., 2021. Climate change and financial risk. Financial risk management and modeling. Springer, pp. 393–419.

Conlon, T., Ding, R., Huan, X., Zhang, Z., 2024. Climate risk and financial stability: evidence from syndicated lending. Eur. J. Financ. 30 (17), 2001–2031.

Curcio, D., Gianfrancesco, I., Vioto, D., 2023. Climate change and financial systemic risk: Evidence from US banks and insurers. J. Financ. Stab. 66, 101132. https://doi.org/10.1016/j.jfs.2023.101132.

Du, K., Xing, F., Mao, R., Cambria, E., 2024. Financial sentiment analysis: Techniques and applications. ACM Comput. Surv. 56 (9), 1–42.

Du, K., Zhao, Y., Mao, R., Xing, F., Cambria, E., 2025. Natural language processing in finance: A survey. Inf. Fusion 115, 102755.

ECB. 2024. Banks must enhance their climate and environmental risk management frameworks (Eurofi Magazine, Issue. ⟨https://www.bankingsupervision.europa.eu/press/interviews/date/2024/html/ssm.in240220~e5cde4c874.en.html#:~:text=We%20need%20to%20push%20banks,clearly%20telling%20us%20that%20the⟩.

Farimani, S.A., Jahan, M.V., Milani Fard, A., 2022. From text representation to financial market prediction: A literature review. Information 13 (10), 466.

Fatouros, G., Soldatos, J., Kouroumali, K., Makridis, G., Kyriazis, D., 2023. Transforming sentiment analysis in the financial domain with ChatGPT. Mach. Learn. Appl. 14, 100508. https://doi.org/10.1016/j.mlwa.2023.100508.

Federal Reserve. 2024. *Pilot Climate Scenario Analysis Exercise* (Summary of Participants' Risk-Management Practices and Estimates, Issue.

Feridun, M., Güngör, H., 2020. Climate-related prudential risks in the banking sector: A review of the emerging regulatory and supervisory practices. Sustainability 12 (13), 5325.

Gentzkow, M., Kelly, B., Taddy, M., 2019. Text as data. J. Econ. Lit. 57 (3), 535–574.

Ghosh, S., Maji, A., Narayana, A., & Naskar, S.K. (2024). IndicFinNLP: Financial Natural Language Processing for Indian Languages. Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024),

Ghosh, S., & Naskar, S.K. (2023). Using Natural Language Processing to Enhance Understandability of Financial Texts. Proceedings of the 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD),

Ginglinger, E., Moreau, Q., 2023. Climate risk and capital structure. Manag. Sci. 69 (12), 7492–7516.

Hosseiny Marani, A., Baumer, E.P., 2023. A review of stability in topic modeling: Metrics for assessing and techniques for improving stability. ACM Comput. Surv. 56 (5), 1–32.

IMF. (2023). *Climate Risks and Financial Stability: What Can Central Banks and Financial Sector Supervisors Do?* (High-level Seminar on Climate Change Issues for Governors and Financial Sector Policymakers, Bangkok, Thailand, Issue. ⟨https://www.imf.org/en/News/Articles/2023/12/12/sp-climate-risks-financial-stability-what-can-central-banks-and-financial-sector-supervisors-do⟩.

Javadi, S., Masum, A.-A., 2021. The impact of climate change on the cost of bank loans. J. Corp. Financ. 69, 102019.

Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 374 (2065), 20150202.

KPMG. (2024). *Banks managing the financial risks of climate and environmental change.* ⟨https://kpmg.com/xx/en/our-insights/esg/banks-managing-the-financial-risks-of-climate-and-environmental-change.html⟩.

Krueger, P., Sautner, Z., Starks, L.T., 2020. The importance of climate risks for institutional investors. Rev. Financ. Stud. 33 (3), 1067–1111.

Lamperti, F., Bosetti, V., Roventini, A., Tavoni, M., Treibich, T., 2021. Three green financial policies to address climate risks. J. Financ. Stab. 54, 100875.

Leippold, M., 2023. Sentiment spin: Attacking financial sentiment with GPT-3. Financ. Res. Lett. 55, 103957. https://doi.org/10.1016/j.frl.2023.103957.

Li, S., Pan, Z., 2022. Climate transition risk and bank performance:Evidence from China. J. Environ. Manag. 323, 116275. https://doi.org/10.1016/j.jenvman.2022.116275.

Liu, X., Liu, J., Hao, Y., 2024. Climate change shocks and credit risk of financial institutions: evidence from China's commercial banks. Emerg. Mark. Financ. Trade 60 (7), 1392–1406.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. arXiv preprint arXiv, 1907. A robustly optimized bert pretraining approach, Roberta, p. *11692*.

Liu, X.-Y., Wang, G., Yang, H., Zha, D., 2023. Fingpt Democr. Internet-Scale data Financ. Large Lang. Models arXiv Prepr. *arXiv:2307.10485*.

Lo, A.W., Singh, M., 2023. From ELIZA to ChatGPT: The Evolution of Natural Language Processing and Financial Applications. J. Portf. Manag. 49 (7).

Lochner, J.H., Stechemesser, A., Wenz, L., 2024. Climate summits and protests have a strong impact on climate change media coverage in Germany. Commun. Earth & Environ. 5 (1), 279.

Maaten, L. v d, Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 (Nov).

Mandel, A., Tiggeloven, T., Lincke, D., Koks, E., Ward, P., Hinkel, J., 2021. Risks on global financial stability induced by climate change: the case of flood risks. Clim. Change 166 (1), 4.

Manning, C.D., 2008. Syngress Publishing. Introd. Inf. Retr.

Milkau, U., 2022. Climate change risk: Demands and expectations imposed on banks. J. Risk Manag. Financ. Inst. 15 (3), 278–300.

Mohit, B., 2014. Named entity recognition. Natural language processing of semitic languages. Springer, pp. 221–245.

Monasterolo, I., 2020. Climate change and the financial system. Annu. Rev. Resour. Econ. 12 (1), 299–320.

Nach, H., 2025. An Analysis of the news coverage of Fintech in Africa: a natural language processing approach. Qual. & Quant. 59, 839–877. Suppl 2.

Nieto, M.J., 2019. Banks, climate risk and financial stability. J. Financ. Regul. Compliance 27 (2), 243–262.

Nocoń, A., 2024. Climate risk in banks–the case of Polish banking sector. J. Mod. Sci. 56 (2), 535–552.
Oyewole, A.T., Adeoye, O.B., Addy, W.A., Okoye, C.C., Ofodile, O.C., Ugochukwu, C.E., 2024. Automating financial reporting with natural language processing: A review and case analysis. World J. Adv. Res. Rev. 21 (3), 575–589.
Pasquini, C., 2020. Climate and Environmental Risk management in Italian banks. BANCARIA 3, 52–61.
Reghezza, A., Altunbas, Y., Marques-Ibanez, D., Rodriguez d'Acri, C., Spaggiari, M., 2022. Do banks fuel climate change? J. Financ. Stab. 62, 101049. https://doi.org/10.1016/j.jfs.2022.101049.
Roncoroni, A., Battiston, S., Escobar-Farfán, L.O., Martinez-Jaramillo, S., 2021. Climate risk and financial stability in the network of banks and investment funds. J. Financ. Stab. 54, 100870.
Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
Takale, D., 2024. Enhancing financial sentiment analysis: A deep dive into natural language processing for market prediction industries. J. Comput. Netw. Virtual 2 (03).
Taylor, N., 2023. Making financial sense of the future': actuaries and the management of climate-related financial risk. N. Political Econ. 28 (1), 57–75.
Uğur, Ö., Kalay, T., Demirel, O., & Yıldırım, S. (2022). Leveraging the power of Natural Language Processing for Financial Intelligence System. 2022 3rd International Informatics and Software Engineering Conference (IISEC),
Upreti, B.R., Back, P.M., Malo, P., Ahlgren, O., Sinha, A., 2019. Knowledge-driven approaches for financial news analytics. Network theory and agent-based modeling in economics and finance. Springer, pp. 375–404.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.
Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564*.
Xiao, J., Wang, J., Bao, W., Deng, T., Bi, S., 2024. Application progress of natural language processing technology in financial research. Financ. Eng. Risk Manag. 7 (3), 155–161.
Yang, S., Li, S., Pan, Z., 2023. Climate transition risk of financial institutions: measurement and response. Appl. Econ. Lett. 30 (17), 2439–2449.
Yinhao, W., 2024. Analyzing the impact of financial news on the stock market using Natural Language Processing techniques. Appl. Comput. Eng. 76 (1), 294–301.
Zhai, P., Fan, Y., Ji, Q., Ma, Y.-R., 2024. Climate risks and financial markets: a review of the literature. Clim. Change Econ. 15 (04), 2440008.
Zhang, Q., Lin, Y., Cao, Y., 2024. From financialization to sustainability: the impact of climate risks on shadow banking activities in non-financial firms in China. Sustainability 16 (19), 8675.
Zhou, F., Endendijk, T., Botzen, W.W., 2023. A review of the financial sector impacts of risks associated with climate change. Annu. Rev. Resour. Econ. 15 (1), 233–256.