Unmasking Climate Risk in Earnings calls: Traversing Storms and Fire with a Taxonomy-GPT prompting approach

Michele Cimino* Annalisa Molino† Maria Paola Priola‡ Lorenzo Prosperi§
Lea Zicchino¶

March 2025

Abstract

Understanding firms' exposure to physical climate risks is crucial as extreme weather events increasingly disrupt supply chains, damage assets, and impact market demand. However, assessing these risks remains challenging due to persistent data gaps, particularly in geolocating critical assets and capturing indirect exposures. To address this, corporate earnings calls serve as a valuable resource for bridging these information gaps. This paper introduces a novel approach using Natural Language Processing (NLP) to extract firm-level climate risk indicators for four major hazards—storms, cold weather, heatwaves, and wildfires—from earnings call transcripts. We develop a taxonomy for climate hazards based on scientific reports to systematically identify relevant climate-related discussions in earnings call transcripts and used GPT-3.5 to analyze corporate exposures. Unlike previous studies, our methodology provides hazard-specific, time-varying metrics that not only distinguish between risks and opportunities but also classify transmission channels as direct or indirect. Our findings show that while climate hazards generally pose risks to firms, some businesses can experience financial gains due to unexpected opportunities arising from extreme weather events. Moreover, we find that the primary transmission channel of physical risk is indirect, impacting firms not through direct damage to assets or production sites but rather via supply chain disruptions and market demand shifts. Finally, our analysis demonstrates that physical risk exposure negatively impacts stock returns, highlighting its significant financial implications across all hazard types.

Keywords: Climate change, Physical risks, Text Analysis, Pattern Matching, Conference calls, Word Embeddings, GPT
JEL: C58 Financial Econometrics, C63 Computational Techniques, G32 Financial Risk and Risk Management, Q54 Climate, Natural Disasters and Their Management, Global Warming

*Dept. of Financial Markets and Intermediaries Analysis, Prometeia. Bologna, Italy. michele.cimino@prometeia.com
†Dept. of Financial Markets and Intermediaries Analysis, Prometeia. Bologna, Italy. annalisa.molino@prometeia.com
‡Dept. of Economics, University of Cagliari, Cagliari, Italy. mariap.priola@unica.it
§Dept. of Financial Markets and Intermediaries Analysis, Prometeia. Bologna, Italy.
University of Edinburgh Business School, B-CCaS lorenzo.prosperi@prometeia.com; Corresponding author
¶Dept. of Financial Markets and Intermediaries Analysis, Prometeia. Bologna, Italy.
Johns Hopkins University SAIS Europe. Bologna, Italy. lea.zicchino@prometeia.com

1 Introduction

In an era of growing environmental uncertainty, understanding firms' exposure to physical climate-related financial risks has become essential. The increasing frequency of climate hazards poses a significant challenge for corporations worldwide. Understanding how these risks impact business is of paramount importance to quantify the effect on the real economy (Batten, 2018; Carleton and Hsiang, 2016; NGFS, 2024) and the transmission to the financial sector (S. Battiston, Dafermos, and I. Monasterolo, 2020; Ranger, Mahul, and Irene Monasterolo, 2022).

These risks can negatively impact the value of corporate assets and infrastructure, cause business interruptions, disrupt supply chains, and alter end-market demand. This suggests that accurately quantifying corporate exposure to physical risk requires detailed information on a firm's structure, assets, and supply chain. Indeed, best practices in physical risk assessment require geolocating not only firm headquarters but also factories and inventories (BCBS, 2021). Neglecting asset-level information may lead to underestimating investors' loss from physical risk (Bressan et al., 2024) and have implications for financial stability as well (Mandel, Stefano Battiston, and Irene Monasterolo, 2025). However, geographic location data is often cited as a major data gap in various climate stress test exercises (FSB, 2022). Moreover, assessing firms' climate-related financial risks requires more than simply mapping their buildings' locations and matching them with geo-located data. While this is an important first step, it does not fully capture the extent of climate risk exposure. Certain facilities within a factory network may be disproportionately critical, amplifying the financial consequences of their exposure to climate hazards. While this information is certainly critical, it cannot be easily retrieved from public sources also for big corporates. Moreover, the impact of these hazards extends beyond direct physical damage to infrastructure. Climate events affecting economically significant regions can have substantial repercussions on businesses, even if the affected firms do not own assets in those locations. Therefore, determining the financial materiality of specific climate hazards for a corporation necessitates extensive data on firm structure, supply chain and end-market demand (Ghadge, Wurtmann, and Seuring, 2019). For large, complex enterprises with numerous subsidiaries, obtaining such granular information is critical yet often infeasible, requiring significant assumptions due to persistent data limitations.

The main contribution of this paper aims to close the existing data gap on firms' material exposure to physical climate risk by leveraging corporate earnings call transcripts to extract firm-level indicators. Earnings call transcripts provide a valuable and underutilized resource for this purpose. As key corporate events, they offer a platform where financial analysts engage with firm management, discussing material developments that may impact the company's financial performance (Hollander, Pronk, and Roelofsen, 2010). This implies that the discussions in earnings calls are inherently prioritized by relevance, as they are time-constrained meetings focused on material financial factors, ensuring that climate risks addressed in these calls are those deemed significant by both firm management and market participants. Moreover, unlike regulatory disclosures, earnings calls provide "soft information" and real-time insights into the risks firms face, including those not explicitly disclosed in public filings.

Building on previous studies that use earnings call transcripts to assess firms' exposure to physical climate risk (Q. Li et al., 2020; Sautner et al., 2023), this paper introduces a novel approach that captures more granular, multi-dimensional indicators of corporate exposure. Leveraging Natural Language Processing (NLP) techniques, we extract time-varying, firm-level metrics that are (i) hazard-specific, covering wildfires, floods hurricanes, cold waves, and heatwaves droughts; (ii) impact-sensitive, distinguishing between risks and opportunities for firm operations; and (iii) channel-aware, differentiating whether the transmission mechanism is direct (e.g., asset damage) or indirect (e.g., supply chain disruptions or demand-side effects). To this end, we define three firm-level indicators: an Exposure Index, based on the frequency of hazard-related references; a Risk Index, capturing the net sentiment by comparing negative and positive impacts; and a Channel Index, which indicates whether the dominant mode of transmission is direct or indirect. We apply these indicators to a large panel of firms to empirically characterize how physical climate risks manifest across sectors and time

By extracting these indicators from a large panel of firms, we uncover how physical climate risks manifest across firms, sectors and time in a heterogenous way. While this heterogeneity may be expected, some findings are particularly striking. For instance, although climate hazards typically pose risks to business operations, a substantial share of firms experience positive exposure, indicating potential financial gains. A clear example of this is the construction industry, where firms can experience windfall profits from post-disaster rebuilding efforts following extreme weather events. Second, our study highlights that the impact of climate hazards on firms is predominantly indirect, often materializing through shocks in end markets or disruptions in supply chains. This underscores the need for a more comprehensive approach

to physical risk assessment—one that extends beyond geolocating firm assets to incorporating detailed insights into firms' value chains and sales. Finally, we highlight the widespread exposure of firms to multiple climate hazards simultaneously, adding complexity to their risk profiles.

Crucially, our second contribution demonstrates the financial materiality of these risk. In particular, we examine whether differences in exposure, risk, and channel explain variations in firms' financial performance using a panel regression approach controlling for firms' characteristics. Importantly, we find that an increase in risk index negatively affects firm valuations, with this effect remaining consistent across all hazard types. To the best of our knowledge, the widespread negative effects of physical risk on stock returns, regardless of the specific hazard considered, have not been previously documented in the literature. Indeed, existing studies has reported inconsistent findings across different hazard types. However, we found that exposure per se does not affect stock returns, suggesting that merely mentioning natural hazards in earnings calls, without addressing their impact, does not influence valuations. Finally, our analysis reveals that the interaction between the Channel index and the Risk or Exposure indexes is not statistically significant, suggesting that financial markets may not differentiate the impact of physical hazards based on the channel of exposure.

These results rely on the third key contribution of the paper: a novel NLP framework, STORM (Semantic Taxonomy and Output Retrieval via prompt-based Modeling), that enables the extraction of climate risk indicators from unstructured financial text. To the best of our knowledge, we are among the first to develop a structured taxonomy framework for identifying climate risk language in corporate disclosures. STORM is a multi-step algorithm that integrates a semantically enriched taxonomy of climate hazards with lexical chaining, distributional semantics, and prompt-based reasoning using large language models (LLMs) to identify relevant climate-related content in earnings calls. Specifically, we construct an initial taxonomy of climate hazards based on institutional sources such as the IPCC and NOAA (Kikstra et al., 2022), and expand it using semantic databases like WordNet and BabelNet. To enhance precision, we combine direct keyword matching with contextual validation, implementing a domain-specific variant of lexical chaining (Morris and Hirst, 1991) that ensures semantic coherence between detected terms and surrounding text. This process is followed by precision-based filtering to remove spurious matches. We then apply a Word2Vec model (Mikolov et al., 2013), trained on the corpus of matched transcripts, to identify additional semantically related terms that refine and expand the taxonomy. The resulting vocabulary enables the extraction of relevant paragraphs from transcripts, which are subsequently classified using a generative AI model (GPT-3.5) (Azure-AI-Services, 2023) to produce structured firm-level indicators of climate risk.

The paper is organized as follows. In section 2, we review the relevant literature. Section 3 describes the main data sources used in the analysis. In section 4, we outline our NLP framework, the STORM algorithm, and how the firm-level indicators are constructed. Section 5 provides descriptive statistics of our indicators across hazards focusing on key differences across firms and sectors. Moreover, we compare our firm-level physical risk indicators with news-based indicators of physical risk (Bua et al., 2022) and available information on major climate events. This section also includes a validation analysis using geolocated data on firms' facilities and major hurricanes in the U.S., along with a detailed breakdown of corporate exposures across sectors and hazards. In section 5.3, we investigate the empirical relationship between sentiment, exposure metrics and stock returns, using a quarterly panel regression analysis. Section 6 concludes.

2 Related Works

2.1 Impact of Physical Climate Risk on Financial Performance

This study contributes to the literature on the impact of physical climate risks on financial performance. Most research focuses on the direct impacts of natural disasters by linking firms' geo-localized assets with hazard data. Donadelli et al. (2020) show that U.S. tornado activity decreases house prices and affects stock returns, especially in the consumer discretionary and telecommunications sectors. Their analysis highlights regional variations, with the South and Midwest experiencing the most significant declines. Kruttli, Roth Tran, and Watugala (2023) use geo-localized data, combining firm establishment and sales information at the county level with hurricane forecast and damage data, to identify firms operating in regions exposed to potential hurricane landfall. This approach allows them to assess market responses to extreme weather uncertainty and its economic impact on financial returns.

Huynh and Xia (2023) leverage disaggregated establishment-level data to examine how investors react

to a firm's exposure to physical climate risk in U.S. bond and stock markets. They find that investors tend to overreact, driving current prices down and future returns up. Similarly, Bourdeau-Brien and Kryzanowski (2017) analyze the effects of natural disasters on U.S. firms, showing that significant impacts on stock returns are limited to firms in disaster zones, with the strongest effects occurring two to three months after peak coverage. Bressan et al. (2024) develop a methodology to assess climate physical risks at the asset level, considering both chronic and acute impacts, such as hurricanes, across IPCC scenarios, highlighting the critical need for granular data in shaping adaptation policies and financial strategies to enhance resilience to climate change. Despite these advances, most of these studies remain confined to direct exposures and require detailed, often unavailable, spatial data on firms' facilities.

A smaller but important strand of the literature focuses on indirect transmission channels. Barrot and Sauvagnat (2016) analyze the impact of natural disasters focusing on upstream effects on the value chain. They found that these effects can lead to significant output and market value losses for customers, with spillover effects for other suppliers. Lenzen et al. (2019) analyze the economic impact of Tropical Cyclone Debbie, which struck northeastern Australia in 2017, using a detailed multi-region input-output framework. They demonstrate that localized production shortfalls caused by storm damage led to widespread job and income losses across upstream supply chains, even in unaffected regions. Although most of the evidence remains aggregated at the sector or regional level, limiting its application for firm-specific financial analysis

The literature also examines the impact of heatwaves and droughts on financial returns. Overall, previous studies indicate that these hazards generally have a negative effect on financial returns, though the magnitude of the impact remains uncertain. Finally, (Addoum, Ng, and Ortiz-Bobea, 2020; Bourdeau-Brien and Kryzanowski, 2017; Choi, Gao, and Jiang, 2020; Hong, Karolyi, and Scheinkman, 2020). Bourdeau-Brien and Kryzanowski, 2017 analyze the impact of fires on stock returns, but they don't find any significant effect.

Our paper contributes to this literature in several key ways. First, we use earnings call data to assess firms' exposure to natural hazards, circumventing challenges associated with geolocating events and linking them to firm subsidiaries while also capturing indirect supply chain effects. Moreover, unlike much of the existing literature, which aggregates natural events or focuses primarily on hurricanes and floods, we distinguish between different types of hazards. Finally, we explore the relative importance of direct versus indirect transmission channels. Compared to existing literature(Barrot and Sauvagnat, 2016; Lenzen et al., 2019), our analysis complements their findings by also capturing downstream effects along the value chain, such as shocks in end-market demand, without relying on sectorial information but using firm-level information extracted from earnings call. This approach enables us to isolate the impact of physical risk on financial returns by controlling for different exposure characteristics (hazard type, risk/opportunity and transmission channel).

Moreover, our findings offer novel insights that complement existing literature on managerial responses to climate disasters, particularly in relation to earnings management. Prior studies, such as Boubaker et al., 2025, suggest that firms may engage in opportunistic earnings manipulation to inflate financial performance in the aftermath of extreme events. Conversely, Byard, Hossain, and Mitra, 2007 show that firms facing heightened public scrutiny—such as petroleum refiners after Hurricanes Katrina and Rita—may adopt conservative reporting strategies, reducing earnings to mitigate political costs. Our analysis adds a new layer to this narrative show that firms also manage qualitative disclosure through the framing of physical climate risks in earnings calls. We find that companies often emphasize indirect channels of exposure or even highlight potential business opportunities arising from climate hazards. These strategic communication choices may serve as a softer form of impression management, particularly when traditional earnings management is constrained by external monitoring or reputational concerns.

2.2 Text-Based Taxonomy Development for Climate Risk

Text analysis has become a key tool for assessing climate risk in finance, offering insights into market perceptions, sector vulnerabilities, and regulatory developments. A growing body of research provides increasingly granular assessments that support more targeted risk mitigation strategies.

Researchers have explored a wide range of textual sources, including news articles (Engle et al. (2020); Ardia et al. (2023); Faccini, Matin, and Skiadopoulos (2023); Bua et al. (2022)), regulatory documents (Berkman, Jona, and Soderstrom (2024), Koelbel et al. (2020)), and firm-specific reports like annual reports, non-financial disclosure reports, 10-Ks, and earnings calls (Sautner et al. (2023), Q. Li et al. (2020), Bingler et al. (2022), Koelbel et al. (2020),Á. I. Moreno and Caminero (2022),A.-I. Moreno and Caminero (2022)).

While some studies focus primarily on transition risks, others expand to include physical risks, employing approaches that range from manual dictionaries to machine learning and taxonomy-based methods. Sautner et al. (2023), for instance, apply a machine learning classifier to detect firm-level exposures to climate opportunities, physical hazards, and regulatory shocks in earnings calls. Q. Li et al. (2020) take a dictionary-based approach to differentiate between physical and transition risks in corporate disclosures. Qiu et al. (2023) employ ontology matching to extract semantic and spatiotemporal insights from geological hazard reports, combining named entity recognition (NER) and gazetteer matching to monitor geohazard dynamics. Oliveira, Mourthe, and Duque (2022) utilize Fuzzy String Matching on drilling reports to identify operational inefficiencies. Wu (2024) leverages a seeded word embedding approach with bag-of-words analysis to measure firm-level supply chain risk from earnings calls, showing strong correlations with option-implied volatility and inventory buffers.

In parallel, semantic enrichment techniques have gained attention for improving classification and interpretation. Ruas, Ferreira, et al. (2020) and Ruas and Grosky (2017) propose lexical chaining methods that combine lexical databases with word embeddings to capture document-level semantic cohesion. These methods have demonstrated improved accuracy in general text classification, though applications in financial climate discourse remain limited.

Our STORM algorithm builds on this foundation but introduces key innovations. First, unlike prior studies that rely on single-source vocabularies or general-purpose taxonomies, we construct a semantically enriched, domain-specific taxonomy tailored to financial discourse around physical climate hazards. Second, while previous research uses lexical chaining for general semantic representation, we operationalize it in a supervised way through a Neighbourhood Matching algorithm with a precision-based threshold, enhancing the accuracy of hazard-specific language detection. Third, we train a Word2Vec model on the matched corpus to identify and integrate semantically similar terms that may not appear in the original taxonomy. This framework allows for the construction of time-varying, granular, and hazard-specific firm-level indicators. Compared to traditional dictionary or sentiment-scoring methods (Loughran and McDonald, 2011), our framework enables context-sensitive interpretation of climate references in short-form financial communication, such as earnings calls.

2.3 Large Language Models for Climate Risk Analysis

Large Language Models (LLMs) have shown broad potential applications in extracting valuable information from unstructured data since Wei et al. (2022). Distilled models like Climate BERT (Webersinke et al., 2021) have shown the potential to outperform LLMs (Bingler et al., 2022; Leippold, 2023). However, these models require significant training efforts, both in terms of human resources, due to the need for large amounts of labeled data, and computational resources, owing to the training of large neural networks. In contrast, large-scale models such as GPTs deliver competitive results without requiring additional refinements or retraining. This advantage is particularly relevant when using prompt engineering to streamline risk extraction tasks. The effectiveness of these models depends on how well they are guided to focus on specific tasks — a process known as prompt engineering. Precise prompt design is crucial for optimizing data retrieval while minimizing irrelevant or inaccurate outputs (Marvin et al., 2023), especially in complex domains like like physical risk extraction from corporate earnings calls.

Both methodologies remain competitive against traditional approaches commonly cited in the literature, such as dictionary-based sentiment indicators (e.g., Loughran and McDonald (2011)) or machine learning algorithms for topic extraction (Bingler et al., 2022). This is particularly relevant when extracting indicators from concise text corpora with limited contextual information. In our dataset, references to physical hazards are often brief and embedded within broader discussions of firm performance,with only a small portion of conference calls devoted to discussing such risks. Generative models like GPT(OpenAI, 2022) can incorporate external information on the context, such as the company's name and primary economic activities, helping to guide data extraction and improve overall performance.

Recent research further underscores the importance of prompt engineering. Jeon, M. Park, S. Park, et al. (2024) demonstrate how GPT models, when carefully guided by prompt logic, can integrate multimodal data to enhance hazard identification in construction environments. Their findings show that by tailoring prompts effectively, GPT can successfully extract key safety risks and hazards from large datasets. Similarly, Collier, Gruss, and Abrahams (2024) explores the use of GPT models in product risk assessment, finding that while LLMs excel in divergent thinking and scenario generation, they struggle with precise domain-specific reasoning without proper guidance. They emphasize that the quality of the generated output is directly tied to the specificity of the prompts, reinforcing that prompt engineering is key to maximizing the utility of LLMs in highly specialized tasks.

Building on these insights, our approach employs GPT-3.5 to extract firm-level physical risk exposure from unstructured earnings call transcripts. The importance of prompt engineering in the STORM algorithm cannot be overstated. We design prompts to guide GPT toward identifying climate-related hazards such as floods, hurricanes, and wildfires, and categorizing the impacts as risk or opportunities as well as direct or indirect. These prompts help GPT focus on extracting hazard-specific information while distinguishing between risks and opportunities. Through iterative prompt refinement, as recommended by Collier, Gruss, and Abrahams (2024), we enhance the model's ability to provide accurate, relevant results. Our work also builds on the lessons from Jeon, M. Park, S. Park, et al. (2024), utilizing prompt logic to optimize GPT's extraction capabilities for climate-related risks in the financial domain. By leveraging these insights, our methodology produces more granular and nuanced risk metrics, offering a clearer picture of how firms are exposed to different climate hazards through both direct and indirect channels.

Our work aligns with and extends existing research by applying advanced prompting techniques to a new domain—physical risks—providing a more comprehensive understanding of firm-level risk exposures.

3 Data

Expanding upon the research of Sautner et al. (2023) we harness publicly available transcripts of corporate earnings calls to extract signals of firms' exposure to physical climate risks. Earnings conference calls serve as a conduit for corporations to relay information and positive guidance to stakeholders and interested parties on the company's financial results. As highlighted by Sautner et al. (2023), earning calls are instrumental for financial analysts and market participants in garnering insights and engaging directly with corporate management. The dynamics of this interaction could significantly mitigate the susceptibility of these communications to corporate greenwashing or non-disclosure in contrast to other climate risk disclosure tools, such as annual reports (Bingler et al. (2022)). Crucially, the transcripts of earning calls typically include not only the presentation of the results by company officials but also a section of questions and answers where officials take direct questions from other call participants.

Earning call transcripts for listed firms are widely available through data platforms or corporate websites. We collect two English-language conference call datasets: the Multimodal Aligned Earnings Conference Call Dataset for Financial Risk Prediction (MAEC J. Li et al. (2020)) and a selection of available transcripts from the LSEG workspace ("Datastream International" 2023). MAEC is a large-scale multi-modal, text-audio paired, earnings-call dataset, based on S&P 1500 companies.

Next, we collect data from LSEG database for 3'144 publicly-listed firms, spanning two decades from 2001 to 2021. Since a bulk download of all available transcripts is unavailable to the authors, we select a sample of firms and manually download all available transcripts. The selection of firms is driven by looking at available climate risk scores developed in Sautner et al. (2023). Specifically, we select all firms with positive physical risk exposures or a non-zero sentiment toward physical risk[1]. To re-balance our set, we randomly select a second set - equal in number to the first exposed set - of firms that are never exposed to physical risks. To this extent, the sample of analyzed firms does not represent the population of globally listed firms; overweight firms are more likely to be exposed to physical and transition risks.

Merging the two databases yields 96'332 transcripts of English-language conference calls. Our dataset includes companies from 19 sectors and with headquarters in 50 countries, although we report a concentration of 68% in North America and 23% in European countries. Nearly 60% of the firms are in the industrial sector, with Manufacturing being the largest represented sector (35%), followed by ICT (14%). The number of firms is initially low, with fewer than 1,000 firms per year before 2004, but stabilizes over time at approximately 1,400 firms annually. Detailed statistics tables are available in the Annex (Section 7.4).

To examine the relationship between physical risk indicators, firms' stock returns, and other firm characteristics, we combine our scores with stock returns and financial statement variables. The data was sourced from LSEG Workspace, and the matching and preparation procedure is detailed in Section 5.3. After excluding illiquid stocks and addressing missing information, the matched dataset used for financial analysis is smaller than the full dataset (10% of firms are excluded), which includes only the exposure indexes.

[1]The physical risk exposure score in Sautner et al. (2023) counts the frequency of specific physical risk bigrams in a transcript, scaled by the number of bigrams. The sentiment score represents a refinement of the previous one by counting the number of climate change bigrams after conditioning on the presence of the positive or negative tone words in Loughran and McDonald, 2011

4 The NLP framework: the STORM algorithm

This section presents the methodological structure of STORM (Semantic Taxonomy and Output Retrieval via prompt-based Modeling), our NLP framework designed to extract structured indicators of firm-level physical climate risk from earnings call transcripts. The algorithm unfolds in three key stages, each detailed in a dedicated subsection. In Section 4.1, we introduce the taxonomy development framework, where we construct a semantically enriched taxonomy of climate hazards by combining expert-defined keywords with lexical resources (WordNet, BabelNet) and expand it through a domain-specific Word2Vec model. We implement a supervised lexical chaining algorithm and apply precision-based filtering to ensure that extracted terms capture climate-relevant discourse with high semantic coherence. Section 4.2 outlines the classification phase, where paragraphs matched via the taxonomy are analyzed using GPT-3.5. Through tailored prompt engineering, the model performs three zero-shot tasks: validating exposure, classifying the impact as a risk or opportunity, and identifying the transmission channel (direct vs. indirect). Finally, Section 4.3 describes how we aggregate these outputs into three firm-level indicators: the Exposure Index, the Risk Index, and the Channel Index, which capture the intensity, polarity, and nature of firms' exposure to physical climate hazards over time.

4.1 A Multi-Step Taxonomy Framework for Climate Hazard Detection

We develop a multi-stage algorithm to generate a taxonomy for four types of climate hazards: heatwaves and droughts, wildfires, storms and floods, and cold waves. Our process begins with the manual compilation of candidate terms—including unigrams, bigrams, and trigrams—drawn from institutional sources. To enhance coverage, we expand this vocabulary using lexical databases such as WordNet and BabelNet, which provide synonyms and semantically related terms that may have been initially overlooked.

To identify climate-relevant content within transcripts, we apply an Extended Boolean retrieval model—a classical method in Information Retrieval (IR) that uses Boolean algebra for structured pattern matching (Lancaster and Fayen, 1973). We implement two matching strategies: Single Key Matching (SKM), which detects precise occurrences of individual taxonomy terms, and Neighbourhood Matching (NHM), which introduces a contextual constraint by requiring co-occurrence of matched terms with "control" keywords within a fixed word window. Control terms include both taxonomy keywords and a set of general climate risk expressions.

Neighbourhood Matching can be interpreted as a domain-specific implementation of lexical chaining—a concept describing semantically cohesive sequences of related terms (Morris and Hirst, 1991), often used in NLP tasks such as summarization and topic segmentation. NHM operationalizes this idea in a supervised setting: it validates ambiguous keywords by examining their proximity to control terms within a 40-word context window, enhancing semantic accuracy in the identification of hazard-related content.

Our taxonomy development follows a three-stage algorithm: **(1) SKM and Precision Filtering.** We use SKM to scan the transcript corpus for initial keyword matches. Each resulting paragraph is manually annotated as a True Positive (TP) or False Positive (FP) based on whether the match correctly signals a climate hazard. We then compute Precision (Raschka, 2014) and classify terms into three groups: (i) keywords with Precision below 50% are discarded, (ii) keywords with Precision equal to or above 90% are retained as lead terms, and (iii) terms in the 50–89% range are considered ambiguous. **(2) NHM Contextual Validation.** We apply NHM to paragraphs not matched in Stage 1, focusing on ambiguous keywords. Each occurrence is validated only if semantically related control terms appear within a 40-word span[2]. Ambiguous terms that exceed an 80% Precision threshold are retained. **(3) Word2Vec Expansion.** To further enhance the taxonomy, we train a Word2Vec model (Mikolov et al., 2013) on a trigram-based corpus built from the previously matched paragraphs. Training is conducted separately for each hazard group. For each retained keyword, we extract semantically similar terms with a cosine similarity above 0.80. These are manually reviewed, and only those relevant to the corresponding hazard category are retained, adding 25 text snippets.

This three-stage process yields a refined, semantically enriched taxonomy of 68 high-precision keywords, which serves as the foundation for extracting firm-level climate risk indicators in the subsequent stages of our NLP pipeline. A detailed workflow is provided in Algorithm 1.

To formalize the algorithmic structure, let $\mathcal{T} = \{t_1, t_2, ..., t_n\}$ represent the corpus of transcripts, with each transcript t_i identified by a unique identifier $id_i \in \mathcal{ID} = \{id_1, id_2, ..., id_n\}$. Let $\mathcal{S} = \{s_1, s_2, ..., s_l\}$

[2]We evaluated multiple context window sizes and selected 40 words based on empirical comparison of Precision metrics.

denote the set of paragraphs extracted from $\mathcal{T}$ using keyword matching. Define the set of hazard event groups as $\mathcal{E} = \{e_1, e_2, ..., e_m\}$, where each e_j corresponds to a specific hazard category.

Each event group e_j is associated with a set of lead keywords $\mathcal{L}_e = \{l_1, l_2, ..., l_p\}$, ambiguous keywords $\mathcal{A}_e = \{a_1, a_2, ..., a_r\}$, and control/context terms $\mathcal{C}_e = \{c_1, c_2, ..., c_q\}$. Let $P \subseteq \mathcal{T}$ represent the subset of transcripts matched through the SKM or NHM process, and $\mathcal{N} = \mathcal{T} \setminus P$ denote the unmatched transcripts.

Let $\mathcal{K} = \{k_1, k_2, ..., k_z\}$ be the final taxonomy. The matching output is stored in a set $\mathcal{R}$, structured by method: - For SKM, each record is a tuple $(id_i, \hat{\mathcal{L}}_i, e_j)$, where $\hat{\mathcal{L}}_i \subseteq \mathcal{L}_e$ is the set of lead terms matched in transcript t_i. - For NHM, tuples are of the form $(id_i, \hat{\mathcal{A}}_i, \hat{\mathcal{C}}_i, e_j)$, where $\hat{\mathcal{A}}_i \subseteq \mathcal{A}_e$ are ambiguous matches and $\hat{\mathcal{C}}_i \subseteq \mathcal{C}_e$ are supporting context terms.

Let $\mathcal{V}$ be the vocabulary extracted from paragraphs in $\mathcal{P}$. The Word2Vec model is defined as a function $\mathcal{W} : \mathcal{V} \to \mathbb{R}^d$ mapping each word to a d-dimensional vector. For any lead keyword $l \in \mathcal{L}_e$, the similarity function $Sim : \mathcal{V} \times \mathcal{V} \to [0, 1]$ measures cosine similarity. The set of semantically similar words for each l is given by:

$$\mathcal{S}_l = \{l' \in \mathcal{V} \mid Sim(l, l') > 0.80\}$$

where only words manually validated for relevance to the hazard category are retained.

Algorithm 1 Algorithm for Transcript Matching and Taxonomy Development

Require: $\mathcal{T} = \{t_1, t_2, \ldots, t_n\}$, the corpus of transcripts with unique IDs $\mathcal{ID} = \{id_1, id_2, \ldots, id_n\}$
Require: $\mathcal{S} = \{s_1, s_2, \ldots, s_l\}$, the set of paragraphs extracted from transcripts
Require: $\mathcal{E} = \{e_1, e_2, \ldots, e_m\}$, the set of climate hazard event categories
Require: $\mathcal{L}, \mathcal{C}, \mathcal{A}$, the sets of lead, context, and ambiguous keywords
Require: $\mathcal{P}, \mathcal{N}$, matched and unmatched transcript subsets
Ensure: $\mathcal{K} = \{k_1, k_2, \ldots, k_z\}, \mathcal{R}$, the final taxonomy and set of matched transcript tuples

procedure STAGE 1: SINGLE KEY MATCHING (SKM)
  Initialize $\mathcal{K} \leftarrow \varnothing$, $\mathcal{P} \leftarrow \varnothing$, $\mathcal{R} \leftarrow \varnothing$, $\mathcal{N} \leftarrow \varnothing$
  for each $t_i \in \mathcal{T}$ **do**
    if $M_{SKM}(t_i, \mathcal{L}) = 1$ **then**
      $\mathcal{P} \leftarrow \mathcal{P} \cup \{t_i\}$
      $\mathcal{R} \leftarrow \mathcal{R} \cup \{(id_i, \hat{\mathcal{L}}_i)\}$
    else
      $\mathcal{N} \leftarrow \mathcal{N} \cup \{t_i\}$
    end if
  end for
  for each keyword $l \in \mathcal{L}$ **do**
    Annotate TPs/FPs and compute Precision(l)
    if Precision(l) $\geq 90\%$ **then**
      $\mathcal{K} \leftarrow \mathcal{K} \cup \{l\}$
    else if $50\% <$ Precision(l) $\leq 89\%$ **then**
      $\mathcal{A} \leftarrow \mathcal{A} \cup \{l\}$
    else
      $\mathcal{L} \leftarrow \mathcal{L} \setminus \{l\}$
    end if
  end for
end procedure
procedure STAGE 2: NEIGHBOURHOOD MATCHING (NHM)
  for each $t_i \in \mathcal{N}$ **do**
    if $M_{NHM}(t_i, \mathcal{A}, \mathcal{C}) = 1$ **then**
      $\mathcal{P} \leftarrow \mathcal{P} \cup \{t_i\}$
      $\mathcal{R} \leftarrow \mathcal{R} \cup \{(id_i, \hat{\mathcal{A}}_i, \hat{\mathcal{C}}_i)\}$
    end if
  end for
  for each keyword $a \in \mathcal{A}$ **do**
    Annotate TPs/FPs and compute Precision(a)
    if Precision(a) $\geq 80\%$ **then**
      $\mathcal{K} \leftarrow \mathcal{K} \cup \{a\}$
    end if
  end for
end procedure
procedure STAGE 3: WORD2VEC SEMANTIC EXPANSION
  Define $\mathcal{W} : \mathcal{V} \to \mathbb{R}^d$ ▷ Word2Vec model trained on corpus from $\mathcal{P}$
  Define $Sim : \mathcal{V} \times \mathcal{V} \to [0, 1]$ ▷ Cosine similarity function
  for each $k \in \mathcal{K}$ **do**
    $\mathcal{S}_k \leftarrow \{k' \in \mathcal{V} \mid Sim(k, k') > 0.80\}$
    for each $k' \in \mathcal{S}_k$ **do**
      Manually verify relevance of k' to hazard category
      if Relevant **then**
        $\mathcal{K} \leftarrow \mathcal{K} \cup \{k'\}$
      end if
    end for
  end for
end procedure

This approach allows us to refine the taxonomy by adapting it to the linguistic patterns used in earnings call transcripts, which often differ from those found in institutional reports. Participants tend to use informal language that may not align with technical or scientific terms. Tailoring the taxonomy

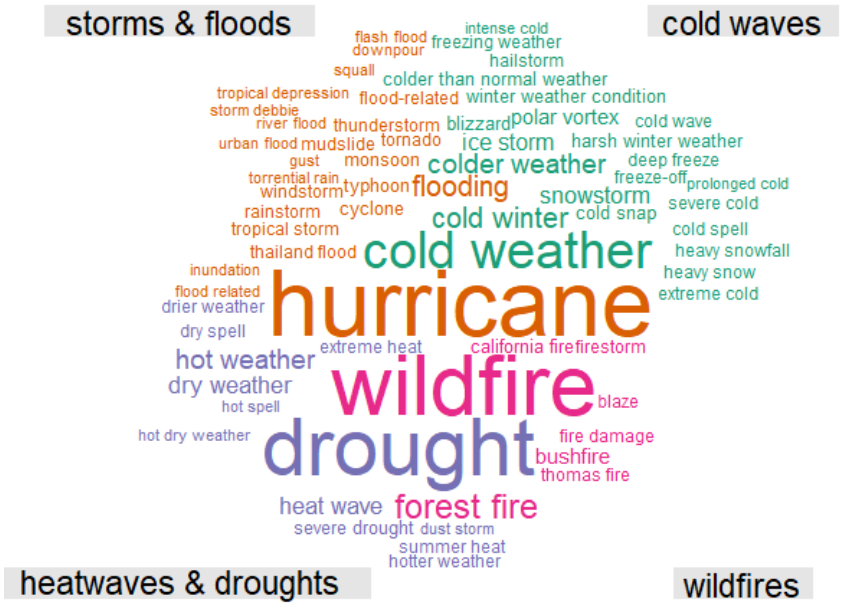


Figure 1: Wordcloud of keywords in the Taxonomy.

This wordcloud illustrates the keywords selected for paragraph retrieval in the proposed approach. The taxonomy was constructed using the Extended Boolean model applied to our set of transcripts, starting from an initial set of hazard-related terms derived from scientific reports. The taxonomy was further enriched using Word2Vec to expand the keyword set. The size of each word in the cloud reflects its frequency of occurrence within our document corpus.

	No. Keywords	No. paragraphs	No. transcripts	No. Firms
Initial taxonomy	43	23301	10488	1771
Word2vec expansion*	25	3939	1821	115
Final taxonomy	68	27240	12309	1886

Table 1: Pattern matching statistics

This table presents statistics on the occurrence of the initial taxonomy, developed using the Extended Boolean model and extension with Word2Vec. The column 'No. paragraphs' indicates the total number of paragraphs identified by all terms within each taxonomy. The final two columns display the number of transcripts and firms associated with the identified taxonomies. Note: * paragraphs, transcripts, and firms not already included in the matched sample.

to this context improves the detection of hazard-related discourse and ensures better alignment with real-world usage. Moreover, by extracting only the paragraphs that explicitly mention climate hazards, our approach narrows the search space for analysis. This targeted extraction is crucial for efficiently deploying Generative AI tools—such as large language models—whose computational costs scale with input length under token-based pricing models. A snapshot of the refined taxonomy is shown in Figure 1, with descriptive statistics reported in Table 1.

Enhancing the taxonomy with Word2Vec leads to a 17% increase in the number of matched paragraphs and a 6.5% increase in the number of exposed firms.

4.2 Prompt Engineering for Physical Risk Identification

We use GPT-3.5, deployed through Azure Machine Learning Studio (Azure-AI-Services, 2023), to analyze paragraphs identified by the taxonomy-based pattern matching (see Section 4.1) and extract structured indicators of corporate physical hazard exposure. The process involves three primary tasks:

First, we validate the precise occurrence of the climate hazard mentioned in the text, ensuring each instance represents a true positive and eliminating any false positives flagged by the Extended Boolean model. This step is crucial for accurately assessing exposure. Second, we classify whether the hazard presents a risk or an opportunity for the firm. When the event threatens the company's operations—such as damage to infrastructure or production disruptions—it is categorized as a "risk." Suppose the hazard

offers a potential business advantage, such as opportunities related to windfall profits from buildings reconstruction or increasing demand of energy use for utilities due to heatwaves, In that case, it is labeled as an "opportunity." If there is no discernible impact or this information is absent, the event is coded as "neutral". As outlined in the appendix in section 7.3, our prompting strategy does not distinguish between hazards that have impacted firms in the past and those anticipated to affect their operations in the future. We regard both types of information as equally relevant. Accordingly, we use the terms "opportunities" and "risks", without explicitly referencing the temporal dimension of exposure.

Finally, we identify the channels through which the hazard affects the corporation's business. The model distinguishes between "direct" impacts, like damage to physical infrastructure, and "indirect" impacts, such as supply chain disruptions or changes in market demand, especially when hazards occur away from the company's core production sites.

These tasks — exposure validation, impact classification, and exposure channels classification — are challenging to identify using conventional methods, especially when hazards affect regions distinct from a firm's main operations. Table 12 provides illustrative examples of these tasks.

	Document	Paragraphs	Labels
Task 1	Q1 2010 Lufkin Industries Earnings Conference Call APRIL 21, 2010 / 2:00PM GMT	"The natural gas market has been a victim of its own success, with shale plays **flooding** the market with more gas than can be concurrently consumed. [...]"	No exposure
	FY 2008 BHP Billiton PLC Earnings Conference Call AUGUST 18, 2008 / 7:30AM GMT	"[...] Metallurgical Coal EBIT was impacted by heavy **flooding** in Queensland in the second half of the financial year."	Exposure
Task 2	Q2 2010 Portland General Electric Company Earnings Conference Call AUGUST 05, 2010 / 3:00PM GMT	"Residential energy sales increased 2.4% due to **colder-than-normal weather** in the second quarter"	Opportunity
	Q2 2021 Natural Grocers By Vitamin Cottage Inc Earnings Call MAY 06, 2021 / 8:30PM GMT	"[...] our second quarter comparable sales were negatively impacted from the **severe cold and ice storms** during February by an estimated 70 basis points."	Risk
	Q3 2007 P&F Industries Earnings Conference Call NOVEMBER 13, 2007 / 4:00PM GMT	"[...] likely to pick up business from the Southern **California fires** and all the new home-building that will have to likely replenish the lots that got burned down."	Indirect
Task 3	Q3 2007 Plum Creek Timber Earnings Conference Call OCTOBER 22, 2007 / 9:00PM GMT	"wildfires damaged approximately 41,000 acres of our forests in the state. Recorded a $4 million loss [...]"	Direct
	Q1 2016 PTT Exploration and Production PCL Analyst Meeting MAY 16, 2016 / 6:00AM GMT	"Another factor which impacts the oil price is supply disruption such as [...] and **wildfire** in Canada, which lowered the supply by around 1 million barrels per day."	Indirect

Table 2: GPT Task and Label Examples

The table presents illustrative examples from the classification and labeling tasks completed by the LLM, drawing from our earning calls' text corpus. Examples are divided by tasks: Task 1 refers to exposure validation, Task 2 refers to impact classification, and Task 3 refers to the exposure channel classification.

To extract these indicators, we craft unique prompts for each task. Refer to the Appendix for additional details on prompt engineering (Section 7.3). We devise three zero-shot prompts, each with explicit instructions detailing: a) the tasks to be executed, including descriptions of the indicators to be generated, b) the desired output format as a JSON document with labeled variables, and c) the analytical context, including a definition of the earnings call, the paragraphs subject to analysis, the hazard keyword, and the company's economic sector by specifying NACE 1-level economic sector classification. While these inputs increase token usage and the overall deployment cost, they significantly enhance performance. To evaluate precision, we manually annotated 1,304 paragraphs for each task and compared human and machine classification. Since Task 1 primarily involves True Positives (with models applied to keywords with 90% and 80% precision), we included 732 out-of-sample negative paragraphs to balance the dataset, sourced from ambiguous text snippets dropped by the Extended Boolean model.

The model's performance is summarized in Table 3. Balanced accuracy is approximately 84% for Task 1, and 71-87% for Tasks 2 and 3. F1 scores are 88% for Task 1, 54-80% for Task 2, and 62-75% for Task 3. These results align with state-of-the-art physical risk metrics using language models for classification tasks (Bingler et al., 2022). We then applied the model to the entire corpus of identified paragraphs, categorizing each by exposure, risk versus opportunity, and exposure channel. Only paragraphs verified by GPT in Task 1 were deemed valid from the initial set identified by the pattern-matching algorithm. We also compare performances in the measurement of the risk and opportunity indicator with two alternative approaches that - similarly to GPT models - do not require any further training to deploy. These approaches are the Loughran-Mc Donald dictionary-based score (Loughran and McDonald, 2011) and a distilled Climate Bert Model (Bingler et al., 2022). We compare results only for the risk vs opportunity indicator because - to our knowledge, this is the only indicator already available in the literature. The comparison confirms that our approach yields satisfactory results when compared to methodologies presented in other studies. Results are discussed in Appendix 7.5.

		Precision	Recall	F1	Balanced Accuracy
Task 1	Class: Exposure	0.91	0.86	0.88	0.84
Task 2	Class: Neutral	0.54	0.55	0.54	0.71
	Class: Risk	0.86	0.78	0.82	0.82
	Class: Opportunity	0.71	0.85	0.77	0.87
Task 3	Class: No channel	0.72	0.79	0.75	0.74
	Class: Direct	0.85	0.66	0.74	0.81
	Class: Indirect	0.62	0.63	0.62	0.73

Table 3: Evaluation results for GPT tasks

The table reports the LLM performance in each classification task. Performances are computed for each labeling class, comparing model outcomes with manual annotations on a subset of paragraphs. Performance indicators are Precision (specificity), Recall (sensitivity), F1 score and Balanced Accuracy. The range goes from 0 to 1 for all scores.

4.3 Firm-Level Indicators of Physical Risk Exposure

The methodology outlined in Section 4.2 is used to extract and classify paragraphs that mention physical risk exposure. This section introduces firm-level indicators of physical risk exposure derived from the classification results.

We define the earnings transcript $E_{i,t}$ of a firm i at time t as a corpus of paragraphs $e_{i,t} \in E_{i,t}$. Each task of the methodology can be represented as an indicator function mapping each paragraph $e_{i,t}$ to the set $\{0, 1\}$. Specifically, we define the following functions:

- **Exposure function**: $g_h(\cdot)$, which takes the value 1 if the paragraph at time t mentions that firm i has been exposed to hazard h (Task 1):

$$g_h(e_{i,t}) = \begin{cases} 1 & \text{if firm } i \text{ is exposed to hazard } h \text{ at time } t \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

- **Risk and opportunity functions**: $r_h(\cdot)$ and $o_h(\cdot)$, which take the value 1 if the paragraph provides evidence of potential harm or opportunity for the firm's economic operations or revenues

(Task 2):

$$r_h(e_{i,t}) = \begin{cases} 1 & \text{if hazard } h \text{ is a risk for firm } i \text{ at time } t \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$o_h(e_{i,t}) = \begin{cases} 1 & \text{if hazard } h \text{ is an opportunity for firm } i \text{ at time } t \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

- **Direct and indirect channel functions**: $d_h(\cdot)$ and $i_h(\cdot)$, which take the value 1 if the paragraph provides evidence of direct or indirect impacts on the firm (Task 3):

$$d_h(e_{i,t}) = \begin{cases} 1 & \text{if hazard } h \text{ impacts the firm } i \text{ directly at time } t \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$i_h(e_{i,t}) = \begin{cases} 1 & \text{if hazard } h \text{ impacts the firm } i \text{ indirectly at time } t \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Using these definitions, we can now define firm-level indicators. Specifically, we define the **exposure index** $\epsilon_{h,i,t}$ of firm i at time t to hazard h as the sum of paragraphs indicating exposure:

$$\epsilon_{h,i,t} = \sum_{e_{i,t} \in E_{i,t}} g_h(e_{i,t}) \tag{6}$$

In Section 5.1, we refer to **exposed firms** at time t to hazard h as those firms with positive exposure, i.e., $\epsilon_{h,i,t} > 0$.

We also define a **Risk index** $\rho_{h,i,t}$ as the difference between the total number of "risk" paragraphs and "opportunity" paragraphs:

$$\rho_{h,i,t} = \sum_{e_{i,t} \in E_{i,t}} r_h(e_{i,t}) - \sum_{e_{i,t} \in E_{i,t}} o_h(e_{i,t}) \tag{7}$$

Similarly, we refer to firms "at risk" as those firms with a positive Risk index, i.e. $\rho_{h,i,t} > 0$, and "opportunity" firms as those with a negative Risk index, i.e. $\rho_{h,i,t} > 0$. Firms with $\rho_{h,i,t} = 0$ are defined as "Neutral". The net risk index could equal zero either because firms do not provide enough information to extrapolate any signal or because the paragraphs providing opposite signals cancel out. Finally, the **Channel index** $\delta_{h,i,t}$ is defined as the difference between the number of directly exposed paragraphs and indirectly exposed paragraphs:

$$\delta_{h,i,t} = \sum_{e_{i,t} \in E_{i,t}} d_h(e_{i,t}) - \sum_{e_{i,t} \in E_{i,t}} i_h(e_{i,t}) \tag{8}$$

In Sections 6 and 5.3, we also utilize aggregate definitions of exposure, risk and channel, which consolidate all hazard-level exposures at the firm level. In particular, they are defined as

$$\epsilon_{i,t} = \sum_{h \in H} \epsilon_{h,i,t} \qquad \rho_{i,t} = \sum_{h \in H} \rho_{h,i,t} \qquad \delta_{i,t} = \sum_{h \in H} \delta_{h,i,t} \tag{9}$$

where H is the set of hazards covered in our analysis. Similarly the aggregate time series of the aforerementioned series can be calculated as follows

$$\epsilon_t = \sum_{t=1}^{T} \epsilon_{i,t} \qquad \rho_i = \sum_{t=1}^{T} \rho_{i,t} \qquad \delta_i = \sum_{t=1}^{T} \delta_{i,t} \tag{10}$$

5 Results

In this section, we present the results of our analysis using firm-level estimates of our physical risk indicators. Section 5.1 includes descriptive statistics on our scores and provides sector-specific insights, highlighting how physical risk exposure, risk and transmission channels vary across different sectors. In Section 5.2, we compare our index at the aggregate level with a news-based index of physical risk and a

Hazard class	Opportunity	Risk	Neutral	Direct	Indirect	No Channel
cold wave	24.31	49.94	25.74	22.44	40.59	36.96
flood&hurricanes&wind	13.29	71.26	15.45	20.94	45.49	33.57
heat wave&drought	27.27	45.11	27.62	11.84	48.20	39.97
wildfires	18.60	57.44	23.97	15.29	28.10	56.61
Total	18.39	66.42	15.20	21.52	51.39	27.09

Table 4: Distribution of Exposure across Hazards, in percentages

The left side of the table shows the distribution (as a percentage of total exposures by hazard) of each hazard categorized as an opportunity, risk, or neutral, based on our earnings call data aggregated at the firm level as described in Section 4.3. Similarly, the right side of the table presents the distribution of exposures across transmission channels—direct, indirect, or no channel

record of major natural hazards from the past two decades. We expect our index to show spikes following major events, reflecting the impact these events have had on firms. Additionally, we conducted a micro-level analysis to identify U.S. firms exposed to hurricanes by matching their geolocations with hurricane paths. Using this methodology, we expect firms identified as directly exposed to hurricanes according to our methodology to be more likely to have facilities located in areas affected by the storms. Finally, in section 5.3 we investigate the relationship between stock returns and our physical risk indicators.

In the following sections, the analysis is conducted on two distinct samples: the "full sample," which includes exposed firms with country and sectoral information[3], and the "restricted sample," comprising firms matched with financial information variables. The two datasets are described in the next section 5.1.

5.1 Descriptive statistics of Physical Risk Indicators

This section aims to analyse the distribution of firms' exposure to the risk metrics described in Section 4.3. Given our population of transcripts described in detail in the appendix 5.1, we extract the physical risk metrics at the firm level as described in section 4.3. The resulting metrics are therefore standardized to have zero mean and unit standard deviation in the full sample. Without integrating these indicators with additional firm-level data beyond NACE sectoral ownership, we obtain a "full dataset" comprising 3,144 firms. Our analysis highlights distinct patterns.

A preliminary assessment of our result can be done by categorizing exposures according to the Risk index and Channel index defined above. As depicted in Table 4, the majority of firms report a negative exposure to hazards. This phenomenon is especially marked in the case of floods and wildfires. Nevertheless, a considerable proportion of firms exhibit positive exposure, particularly following cold waves and heat waves. This suggests that a significant number of firms have referenced specific hazards in their earnings calls, indicating that these events have been (or could potentially be) opportunities for their business. Importantly table 4 provide also the distribution of exposures between different channels. The most striking result is that firms are more likely to be indirectly rather than directly exposed to physical hazards. While there is more than half of the firms which is indirectly exposed to the hazard, there is also a proportion of firms that are directly affected, especially through cold waves and floods.

In Table 18 of the appendix 7.4, we provide evidence of the correlation between different hazards and metrics. On average, the correlation of our risk metrics across hazards are generally positive. More specifically, the exposure indices are positively and significantly correlated across hazards, highlighting that firms often face multiple exposures to different hazards. Indeed, identifying firms exposed to each hazard as those with $\epsilon_{h,i,t} > 0$ reveals that nearly 30% of all firms are exposed to at least two hazards. Focusing on the correlation between risk indices, the correlations appear weaker compared to those calculated for the exposure indices. Specifically, only the pairs flood and cold waves, and cold waves and wildfires, show significant positive correlations. This could be attributed to variations in the 'sign' of exposures among firms, as some are negatively affected while others benefit from the hazards, as shown in Table 4.

An assessment of sector-specific vulnerability (refer to Figure 2) indicates a pronounced disparity in hazard exposure among sectors. Flood-related events constitute the primary risk across all sectors, followed by cold waves as the secondary hazard. Heatwaves and wildfires present a more varied pattern of exposure. Specifically, the agricultural sector (NACE A) and water-related industries (NACE E) are

[3]The population of transcripts available in the full sample is described in 7.4

comparatively more susceptible to heatwaves and droughts. In opposition, the construction industry faces a heightened risk from wildfires.

The second analysis we conduct aims to examine the risk and opportunity patterns associated with each hazard across various sectors. Looking at the diversification by economic sector could provide some intuitions in explaining this tendency in our data. Figure 3 portrays the distribution of risk and opportunity across sectors for different hazards. It becomes apparent that certain sectors face relatively lower exposure to specific hazards, whereas others might even reap advantages. For example, firms engaged in electric power generation (NACE D) are intuitively positioned to gain from cold waves, as demand for utilities tends to rise rapidly during such events.

Next, we investigated which sectors are impacted directly through damage to proprietary assets or infrastructure and disruptions of operations, and which are impacted indirectly, through disruptions to the supply chain or the demand of products. It is clear from Table 4 that firms are more likely to be indirectly rather than directly exposed to physical hazards. The results of the analysis at the sectorial level are shown in Figure 4. The mining sector stands out because it is directly exposed to all risks, a result somehow expected. The agricultural sector tends more often to be exposed directly rather than indirectly, while among the most indirectly exposed sectors, we find Wholesale and retail trade and Accommodation and food service activities.

We then construct a firm-level dataset by combining our physical risk metrics with financial statement indicators and stock returns. Compared to the initial sample, 284 firms (approximately 9%) were excluded due to missing time series data on returns and balance sheet indicators. Specifically, balance sheet data were sourced from LSEG Workspace and processed by retaining only end-of-year values, with missing entries imputed using MICE algorithm (Van Buuren and Oudshoorn, 2000). The selection of balance sheet indicators aligns with relevant literature, as detailed in sections 2 and 5.2. All variables, except for the logarithm of total assets, were winsorized at the 1% level. The dataset includes statistics for the exposure index (Exposure), the risk index (Risk), and the channel index (Channel). As described above, these indexes are standardized to have a mean of zero and a standard deviation of one across the full sample. For the econometric analysis of stock returns in Section 5.2, we collected daily stock price data from LSEG Workspace to calculate cumulative quarterly returns. Illiquid stocks were excluded, and returns were winsorized at the 1% level. Summary statistics for the final sample are provided in Table 17 in the appendix 7.4.

The restricted sample provides valuable insights into the correlation between firms' exposure to physical risk and their characteristics. To this end, in Appendix 7.6, we analyze how economy-wide, industry-specific, and firm-level factors contribute to our newly developed physical risk measures. Regression results examining the relationship between firm characteristics and the two physical risk indices indicate that larger firms and those with a higher share of tangible assets tend to be more exposed to climate hazards. Furthermore, variance decomposition analysis shows that both firm-specific and industry-level factors play a central role in determining corporate exposure to physical risks, emphasizing the significance of sectoral and firm-level vulnerabilities in shaping risk dynamics. Compared to previous studies (Q. Li et al., 2020; Sautner et al., 2023), our indicators suggest that physical risk exposure is more strongly influenced by industry-wide factors than previously documented, highlighting the need to account for sectoral characteristics in climate risk assessments.

5.2 Benchmarking Physical Risk Exposure Indicators Against Alternative Measures

5.2.1 Comparison of the aggregate Exposure Indicators with Global Hazard events and global news index

In this section, we aim to provide an analysis of indicators derived at the firm level. Our initial focus is on evaluating whether the total number of exposed firms in our sample to a specific hazard tends to be higher in years coinciding with significant events. To accomplish this, we conduct a preliminary assessment as follows. After determining the individual exposure of each firm, we track the number of exposed firms across different hazards over time. Subsequently, we compile a list of major global events. The outcomes of this analysis are depicted in Figure 5. The data reveals an increase in the number of exposed firms to hurricanes following the years in which the considered hurricanes struck the US. A similar trend is observed for the other risks. These findings provide preliminary evidence of a consistent correlation between exposed firms and the incidence of major physical events.

We conduct another graphical analysis by comparing the number of exposed firms with a news-based

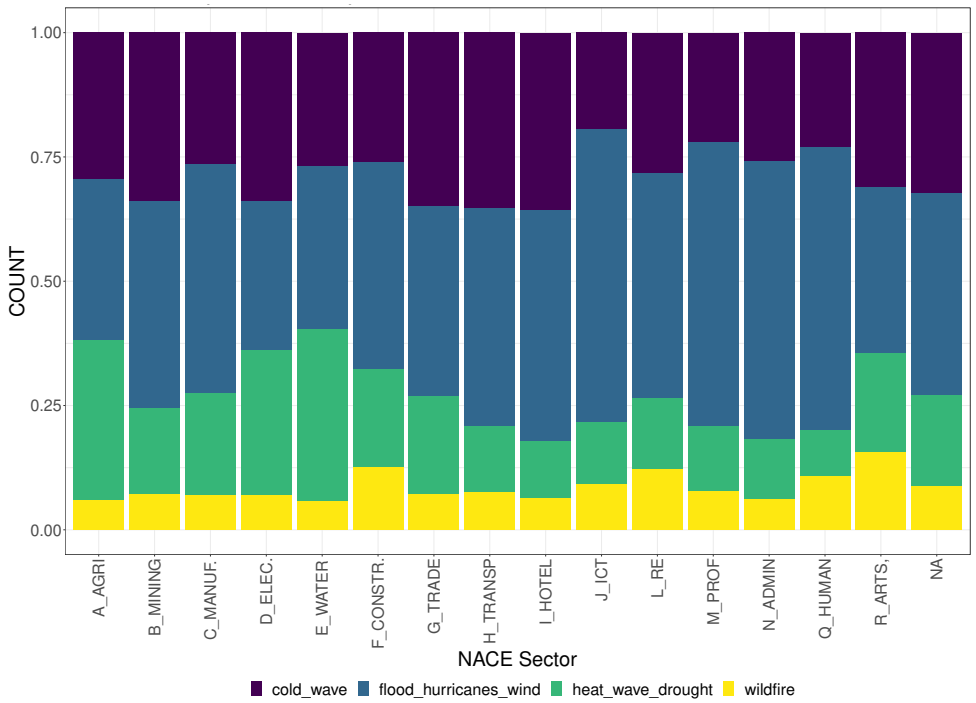


Figure 2: Distribution of Hazard Exposure Across Sectors

For each sector, we count the number of firms exposed to each hazard. Exposed firms are those that mention a specific hazard impacting their business in their earnings transcript. The figure shows the proportion of exposures attributed to each hazard for every sector. Since a single firm can be exposed to multiple hazards, the total number of exposures exceeds the number of exposed firms within a sector.

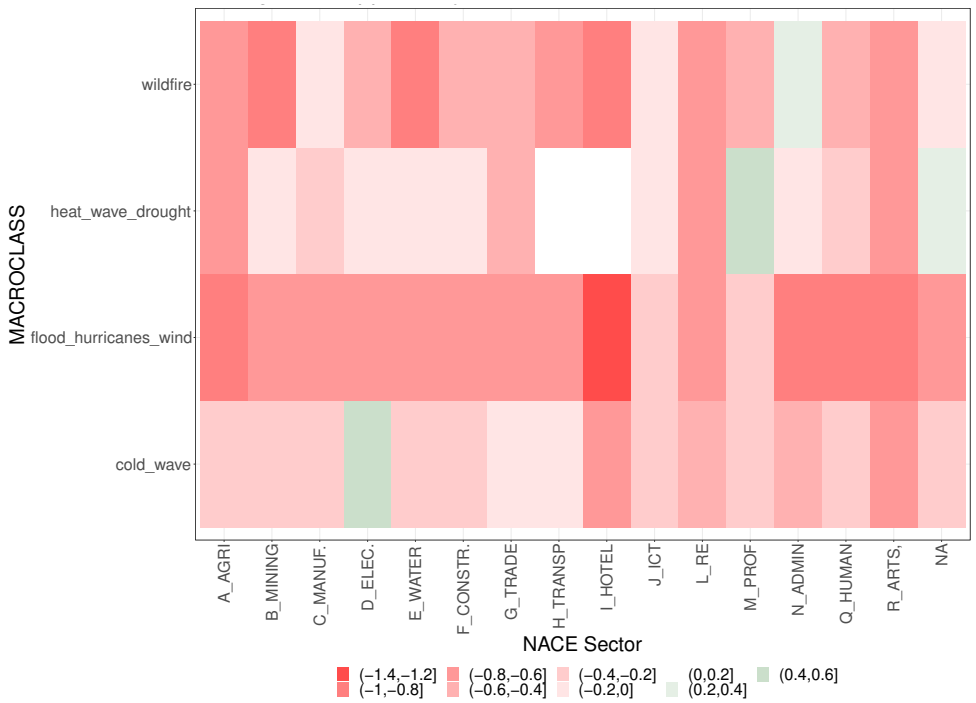


Figure 3: Average Risk Index Across Sectors and Hazard Classes

We compute the risk index as the difference between the total number of "risk" paragraphs and "opportunity" paragraphs for each firm. The figure shows the average risk index for every sector and hazard. Darker tones of red (green) map into a more negative (positive) risk index, meaning more "at risk" ("opportunity") firms for the sector.

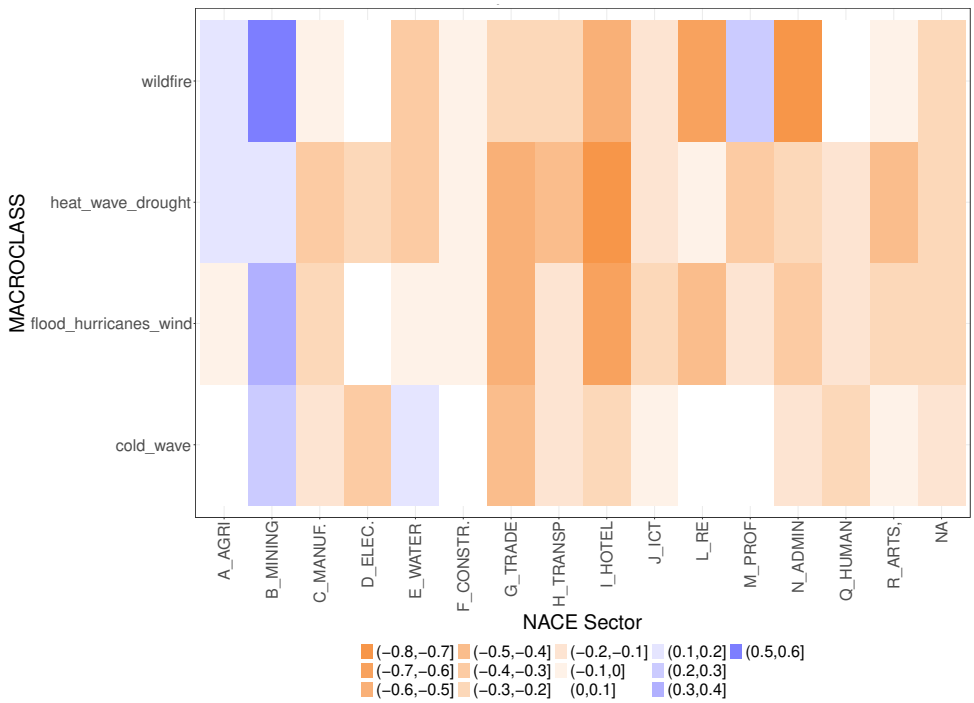


Figure 4: Average Channel index across sectors and hazard classes

We compute the channel index as the difference between the total number of "directly exposed" paragraphs and "indirectly exposed" paragraphs for each firm. The figure shows the average channel index for every sector and hazard. Darker tones of blue (orange) map into a more positive(negative) channel index, meaning more firms "directly" ("indirectly") exposed in the sector.

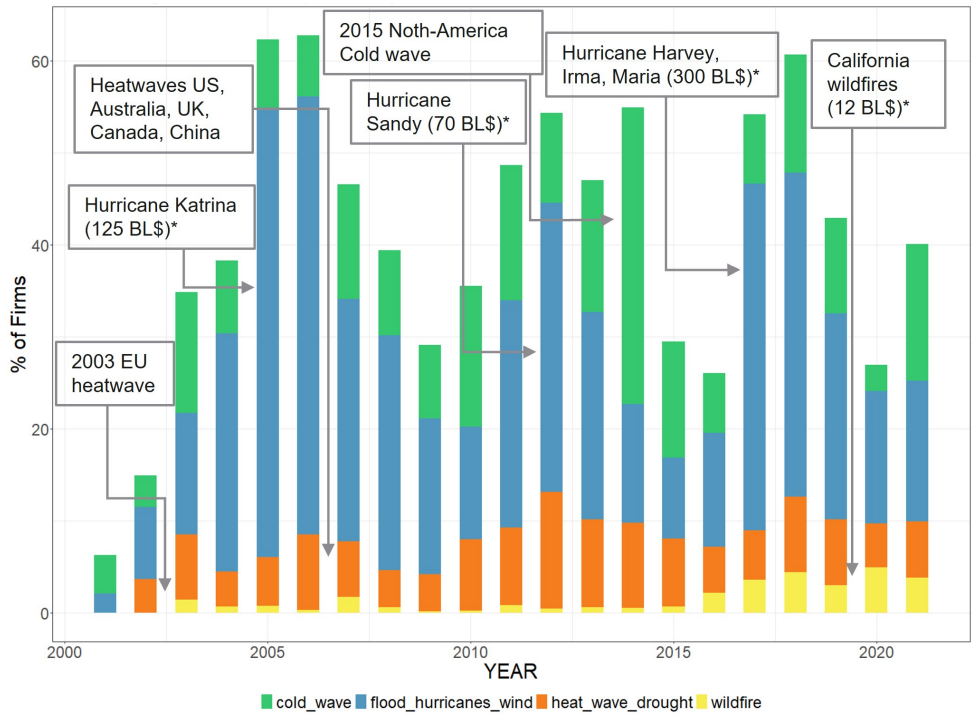


Figure 5: Percentage of firms exposed to hazard by year and hazard

The figure shows the percentage of firms exposed to a hazard in each year of the analysis. The main spikes observed in the indicator coincide with major hazards that occurred in the period.

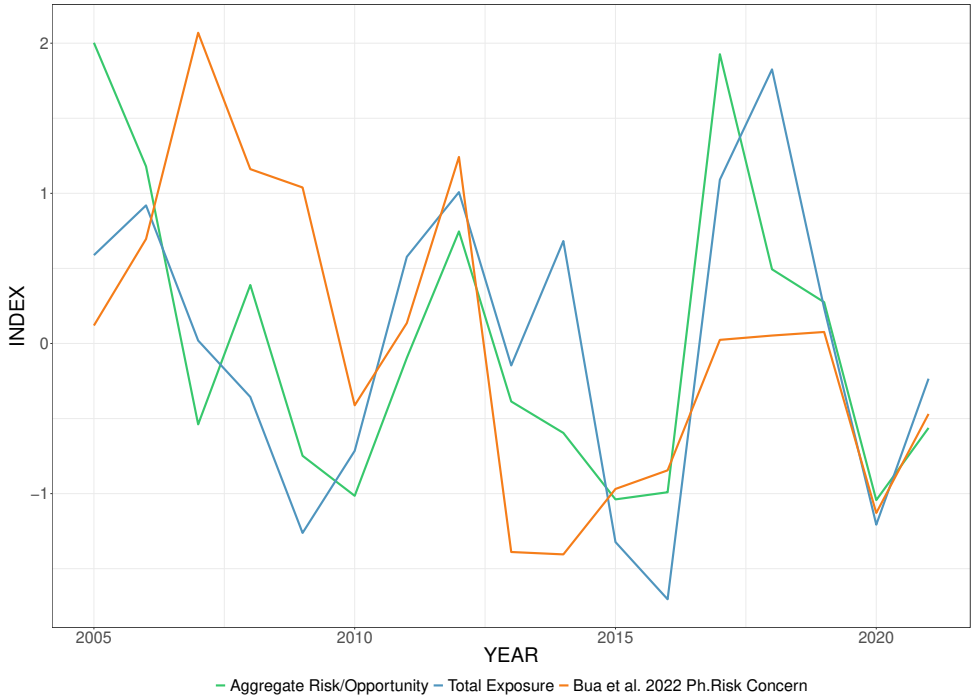


Figure 6: Bua et al. 2022 Climate Risk Concern Index (PRI) vs Aggregate Exposure and Aggregate Risk indicator

The figure shows yearly averages of the time series of Physical Risk Index (Bua et al., 2022) compared with aggregate exposure and risk index calculated as specified in section 4.3. The aggregate time series have been standardized.

aggregate Physical Risk Indicator (PRI) developed in Bua et al. (2022). The PRI is derived from textual analysis of global news articles and aims at capturing news attention over climate physical risks. Our objective is to assess whether the number of exposed firms increases during periods when the PRI spikes, indicating a substantial rise in discussions about hazards, potentially due to significant events occurring. Furthermore, we compare Bua's original indicator with our aggregate risk/opportunity indicator. Both measures are standardized and aggregated on a yearly basis before comparison. The results are illustrated in Figure 6. Although there are substantial discrepancies in the initial portion of the sample, we observe a degree of co-movement between the two measures. Notably, the PRI peaked in 2012 and again during 2017-2019, consistent with our findings. In particular, the correlation between our measures and the PRI is 0.26 when considering the exposure index, and 0.36 considering the risk index.

5.2.2 Comparison of firm-level Exposure indicators with Geolocated Data on U.S. Hurricanes

We further deepen our validation analysis by introducing geo-localized data on the company's assets and hazards. We aim to match spatial data of companies' headquarters, subsidiaries and factories with a precise geo-localization of climate hazards, deriving an exposure index which is completely independent of our climate risk exposure metrics. This analysis is only available for US-based firms and Hurricane events, due to limitations in data accessibility for other geographical regions and hazard types.

The methodology includes the following steps. We consider a restricted sample of firms from our dataset that are headquartered in the US, dividing our sample between firms negatively exposed to Hurricanes, floods and storms and firms not exposed to such hazards. We obtain firm headquarters location data from the LSEG workspace for both the parent company headquarters and the headquarters of their related organizations (primarily subsidiaries). We also incorporate factory locations using the US Environmental Protection Agency (EPA) Program's Facility Level Information on Greenhouse Gases Tool (FLIGHT) ("Greenhouse Gas Reporting Program (GHGRP)" 2023), which compiles location data for factories reporting direct greenhouse gas (GHG) emissions along with the names of their parent companies[4]. Spatial data is provided in different meanings, including latitude and longitude coordinates,

[4]Since October 2009, the EPA has managed the Greenhouse Gas Reporting Program (GHGRP), which publishes data on

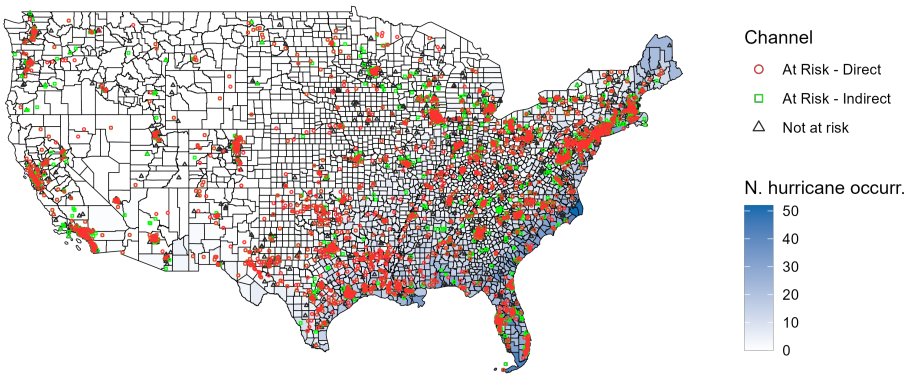


Figure 7: Hurricane occurrences by county and geolocalization of firm's headquarters, subsidiaries and factories.

The figure shows the geolocalized firm's headquarters, subsidiaries, and factories, colored by the firm's exposure channel. Counties are colored based on the total occurrence of hurricanes over the period (2000-2021).

postal codes, and city and state pairings. We match the different coordinates to the location of every firm facility by county and state, which will be the spatial unit of analysis for our validation exercise. Focusing exclusively on acute climate hazards with a consistently large area of effect such as hurricanes, ensures a high degree of confidence that the matches represent significant disruption to firm operations in impacted counties.

We then collect spatial data on Hurricanes from the NOAA (National Oceanic and Atmospheric Administration) both Pacific and Atlantic hurricane tracking datasets ("HURDAT 2", Landsea and Franklin, 2013). These datasets provide latitude and longitude tracking coordinates, which are recorded whenever the hurricane changes direction. This level of information allows us to recreate the exact path of each hurricane, from landfall to dissipation, and to identify every county through which the hurricane has passed. The latter step is completed through the merging of hurricanes' tracking coordinates provided by NOAA and the US Census Bureau's Topologically Integrated Geographic Encoding and Referencing System ("TIGER" 2023) county database — which treats each county as a multi-polygon geometry variable. It is important to note that, since the hurricane's tracking data only provides the path for the hurricane's epicentre, the consequent exposure measurement might underestimate the real hazard impact, as nearby counties might also be affected. A visual map of hurricane and firms' geolocalised assets is depicted in 7.

We finally match the hurricane tracking data with our climate hazard exposure data using both facilities' locations and the year-quarter when earning calls are held. In this way, we can compute the share of firms that mention a negative exposure to hurricanes (referred to as "At Risk" in table 5 and 6) in a conference call and that have at least one facility in a county where a hurricane has passed at the same time of the call. As we might assume some delay from the event occurrence and the firms' disclosure of exposure during an earning conference call, we also allow for additional flexibility in the timing dimensions. We settle for a matching that allows for a delay in the exposure disclosure of up to two years, but results are consistent even with shorter time windows. The same matching exercise is conducted with non-exposed firms.

scope 1 emissions from all U.S. establishments that emit 25,000 tons or more of CO2 equivalents each year. EPA data has been used in Benincasa et al., 2024 to analyze how climate transition risks can significantly disrupt supply chain networks.

	Exposure	Channel	N. Firms	Average N. Facilities per firm
Total Number of Firms	At Risk	Direct	308	12.8
		Indirect	461	14.4
	Not At Risk		1117	7.0
Number of Matched Firms	At Risk	Direct	191	3.7
		Indirect	238	3.6
	Not At Risk		438	1.8
% of Matched Firms	At Risk	Direct	62.0	-
		Indirect	51.6	-
	Not At Risk		39.2	-

Table 5: Exposure to flood and hurricane across total and "matched" US-headquartered firms

We define "matched firms" as U.S.-headquartered companies with facilities located in U.S. counties impacted by cyclones, as identified in the NOAA dataset. Within this group, we further identify firms directly and negatively exposed to floods and hurricanes, based on earnings call transcripts. Our analysis confirms that the proportion of firms identified as directly and negatively exposed is higher among matched firms.

	Exposure	Channel	NACE Economic Sector											
			B	C	D	F	G	H	I	J	L	M	N	Other
Tot. N. Firms	At Risk	Direct	52	83	16	13	24	27	16	15	16	3	9	34
		Indirect	25	158	18	20	58	31	25	27	19	13	19	48
	Not At Risk		50	429	12	14	68	20	8	268	38	122	20	68
N. Matched Firms	At Risk	Direct	33	46	10	12	13	19	9	9	9	1	7	23
		Indirect	15	74	9	15	31	20	12	10	9	8	8	27
	Not At Risk		22	153	9	7	29	9	4	113	11	45	6	30
% Matched Firms	At Risk	Direct	64	55	63	92	54	70	56	60	56	33	78	68
		Indirect	60	47	50	75	53	65	48	37	47	62	42	56
	Not At Risk		44	36	75	50	47	45	50	42	29	37	30	44

Table 6: Exposure to flood and hurricane across total and "matched" US-headquartered firms by economic sector

We define "matched" firms as U.S.-headquartered companies with facilities located in U.S. counties impacted by cyclones, as identified in the NOAA dataset. Within this group, we further identify firms directly and negatively exposed to floods and hurricanes, based on earnings call transcripts. Our analysis confirms that the proportion of firms identified as directly and negatively exposed is higher among matched firms.

We depict the results of our validation exercise in Table 5 and 6, reporting the total number and share of matched firms and facilities. Results are shown by NACE-1 economic sectors and, most importantly, by the direct and indirect exposure channels identified in our analysis.

Consistent with our expectations, we observe higher matching rates between the two metrics in the directly affected firms sample. We expect less correlation with indirectly exposed firms as the disclosed hazard exposure is caused by shocks in the firm's supply chain or end markets, whose geographical locations might not coincide with the location of firms' facilities. Results are also consistent across NACE-1 economic sectors[5]. Firms that do not report exposure to hurricanes in their earning conference

[5]We aggregate NACE-1 sectors with less than 45 firms in a unique category "Other"

calls also have a lower but non-zero matching rate with external data. It is consistent with our analysis to have few matches even in this category, as a simple exposure might not imply a relevant economic impact for the firm and a consequent mention in an earning call.

5.3 Impact of Physical Risk Exposure on Stock Returns

In this section, we investigate whether differential values in the exposure, risk and channel indexes can account for variation in firms' financial performances at the firm level. First, we aim to determine whether exposed firms exhibit different stock returns compared to non-exposed firms and how these differences vary across different types of hazards. A significant relationship between the exposure index and the stock returns may suggest that exposure to physical hazards changes investors' expectations of future cash flows or discount factors. In principle, if the exposure index is positive, the impact on returns can be positive or negative depending on whether the exposure to the hazard is a risk or opportunity for the firm. However, we expect that if the exposure provided a potential harm to the company's economic operations or revenues then it should be associated with lower stock returns (vice versa in case of opportunity). Therefore we expect a negative relationship between the risk index and the firm's financial performance. Second, we investigate the relationship between the stock returns and the channel through which the events impact the company by studying the interactions of the exposure and/or risk index with the channel index. Our goal is also to understand whether our results are driven by specific hazards.

To analyse the relationship between our measures and firms' financial performances we run an econometric analysis to isolate the effects of the indexes in explaining firms' stock returns. To run the analysis we use panel data regression, a standard practice in the literature when investigating the relationship between firm-level exposure measured by textual analysis of corporate earning calls on real and financial outcomes. Sautner et al. (2023) run panel regression to examine whether the levels of their climate textual measures are useful to predict real outcomes related to the net-zero transition (job creation in disruptive green technologies and green patenting) and that they can be associated with changes in financial variables related to the option and equity markets. Q. Li et al. (2020) use panel regression to examine whether their proxy of exposure to transition and physical risk extracted from conference call transcripts are related to balance sheet variables and whether they are reflected in firms' future real outcomes (Investments, Innovation and Employment) and valuations (Tobin's Q). Similarly, in their seminal paper, Bolton and Kacperczyk, 2021 employ a comparable panel model framework to examine the relationship between CO2 emissions and stock returns, revealing that more polluting firms tend to achieve higher returns. In a similar fashion, other authors constructed firm-level measures of exposure and risk using earning conference calls and ran panel regression to investigate whether their exposure and/or risk measures could account for variation in the firm's outcome. Hassan et al. (2019) and Hassan et al. (2020a) in the context of geopolitical risk, Hassan et al. (2020b) for risks related to the spread of epidemic diseases (e.g. COVID19), and Jamilov, Rey, and Tahoun (2021) for cyber risk.

We focus on the estimation of the following type of panel regression using quarterly data.

$$r_{i,t} = \beta_0 + \beta_1 \epsilon_{i,t} + \beta_2 \rho_{i,t} + \Gamma X_{i,t} + \varphi_t + \gamma_i + \chi_{s,t} + u_{i,t} \tag{11}$$

where $\epsilon_{i,t}$ is the exposure index of firm i at time t, and $\rho_{i,t}$ is the Risk index of firm i at time t. We include a vector of balance sheet indicators $X_{i,t}$ as control variables, including the logarithm of total assets, sales or revenue growth, and total debt, cash, CapEx, EBIT, Property, Plant & Equipment (PPE), Research & Development expenses (R&D) divided by total assets[6]. We also include a time effect φ_t (year-quarter), firm-specific fixed effects γ_i and NACE4-year-quarter fixed effects $\chi_{s,t}$ to account for time-varying heterogeneity across industries. In the estimation, we always use clustered standard errors at the firm level.

To study the interaction with the channel of exposure, we extend the model as follows:

$$r_{i,t} = \beta_0 + \beta_1 \epsilon_{i,t} \times \delta_{i,t} + \beta_2 \rho_{i,t} \times \delta_{i,t} + \Gamma X_{i,t} + \varphi_t + \gamma_i + \chi_{s,t} + u_{i,t} \tag{12}$$

where $\epsilon_{i,t}$ is the exposure index of firm i at time t, $\rho_{i,t}$ is the risk index of firm i at time t, and $\delta_{i,t}$ is the channel index.

[6] We test an alternative specification by lagging balance sheet variables by one period to mitigate potential contemporaneous effects of natural hazards on these indicators, which could distort the estimated coefficients of physical risk metrics on stock returns. The coefficients for the physical risk scores remain consistent in magnitude, and the detailed results are available upon request.

The model is estimated on the restricted dataset matching stock returns, firms' physical score and balance sheet indicators described in section 5.1.

Table 7 shows the results of the baseline regression (11). Three different specifications are considered: Model 1 regresses the stock returns on the exposure index, Model 2 regresses the returns on the risk index, and Model 3 includes both indexes as regressors. [7] Only in Model 1, we find a negative and statistically significant effect of the exposure index in the full sample. Specifically, one standard deviation of the exposure risk index above its mean leads to a 0.1% drop in stock returns in that quarter. In Models 2 and 3 we find that the Risk coefficient is negative and statistically significant, consistent with the fact that if exposure to climate hazard is a potential harm to the firm, it could lower its market valuation. This result also suggests that the risk index is the primary determinant of stock returns. In a similar fashion, one standard deviation of the risk index above its mean leads to a 0.1% drop in stock returns in that quarter. This coefficient captures the negative effect unrelated to the firm's specific features, which are accounted for by the fixed effects.

	Model 1	Model 2	Model 3
Exposure	-0.11^{**}		-0.07
	(0.04)		(0.04)
Risk		-0.17^{***}	-0.15^{***}
		(0.04)	(0.04)
Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Industry x Time FE	Yes	Yes	Yes
Num. obs.	136301	136301	136301
Num. ISIN	2826	2826	2826
R^2 (full model)	0.44	0.44	0.44
Adj. R^2 (full model)	0.29	0.29	0.29

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7: Stock Returns and Aggregate Physical Risk Exposure: Estimation Results for Model 11

The table presents the estimation results of the model examining the relationship between financial returns and physical risk metrics—specifically, the aggregate exposure and risk indices derived from earnings calls. The model is estimated at quarterly frequency and includes several balance sheet indicators as control variables. The analysis is conducted on a restricted sample of firms where stock returns, physical risk metrics, and balance sheet indicators are matched. Firm, time, and firm-time fixed effects are incorporated into the model, and standard errors are clustered at the firm level

Table 8 shows the same results by hazard. We find that the negative effect on stock returns is stronger and more significant for firms exposed to floods and hurricanes. However, the coefficient of the risk index is negative and statistically significant across hazards. These results could be explained by the nature of these events, specifically floods and hurricanes, which cause substantial economic damage and disrupt operations, leading to stronger effects on stock returns. Additionally, the geographical location of the companies in the sample and the sector of operations could play a role.

The existence of a negative effect of floods and hurricanes and stock returns is consistent with the findings in the literature (Bourdeau-Brien and Kryzanowski, 2017; Donadelli et al., 2020; Horváth et al., 2024; Kruttli, Roth Tran, and Watugala, 2023). Similarly, other papers have shown that heatwaves, extreme temperatures and droughts are negatively correlated to stock returns however the magnitude of the effect is uncertain (Addoum, Ng, and Ortiz-Bobea, 2020; Bortolan, Dey, and Taschini, 2024). Our findings suggest that also negative exposure to wildfires are also associated to lower stock return, in contrast with previous findings that have not found significant effects Bourdeau-Brien and Kryzanowski, 2017.

Table 9 presents the results of the regression with interaction terms (12). We define three specifications to consider the effect of the interaction between the channel of exposure and our existing regressors: Model 1 includes the interaction between the channel index and the risk index (omitting total exposure), and Model 2 includes the interaction between the channel index and exposure, also entering the risk index as an

[7] We performed a Variance Inflation Factor (VIF) analysis to check for potential multicollinearity among the independent variables. The results indicate that all VIF values are below 2, suggesting that multicollinearity is not a concern in our model. Results are available upon request.

	cold wave			heat wave&drought			flood,hurricanes&wind			wildfire		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Exposure	−0.05		−0.05	−0.03		−0.01	−0.11**		−0.08	0.03		0.07
	(0.04)		(0.04)	(0.04)		(0.04)	(0.04)		(0.04)	(0.04)		(0.05)
Risk		−0.07*	−0.07*		−0.08*	−0.07*		−0.12**	−0.10*		−0.07*	−0.10*
		(0.03)	(0.03)		(0.03)	(0.03)		(0.04)	(0.04)		(0.03)	(0.04)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry x Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	136301	136301	136301	136301	136301	136301	136301	136301	136301	136301	136301	136301
Num. ISIN	2826	2826	2826	2826	2826	2826	2826	2826	2826	2826	2826	2826
R^2 (full model)	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
Adj. R^2 (full model)	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 8: Stock Returns and Hazard-level Physical Risk Exposure: Estimation Results for Model 11

The table presents the estimation results of the model examining the relationship between financial returns and physical risk metrics—specifically, the hazard-level exposure and risk indices derived from earnings calls. The model is estimated at quarterly frequency and includes several balance sheet indicators as control variables. The analysis is conducted on a restricted sample of firms where stock returns, physical risk metrics, and balance sheet indicators are matched. Firm, time, and firm-time fixed effects are incorporated into the model, and standard errors are clustered at the firm level.

additional independent variable. Finally, Model 3 includes both the interaction between the net channel index and the risk index, and the interaction between the net channel index and the exposure index. The coefficient for Risk is consistently negative and statistically significant across all three models. Neither the coefficient for Channel nor the interactions with Risk and Exposure are statistically significant across models. This is in contrast with Barrot and Sauvagnat (2016) that find that firms' returns experience an abnormal drop when one of its suppliers is hit by a major natural disaster. This suggests that a different channel of exposure as measured by our index doesn't play a role when assessing the impact of physical risk on stock returns. These findings suggest that financial markets may not differentiate the impact of natural hazards based on whether a firm is directly or indirectly exposed. This could indicate that the information available from firms' earnings calls is insufficient for markets to make such distinctions.[8]

[8]We explored alternative model specifications by replacing the channel index with two dummy variables to differentiate exposure types. INDIRECT is set to 1 if the channel index falls below the 25th percentile of its historical distribution, while DIRECT is set to 1 if it exceeds the 75th percentile. The coefficients for physical risk scores remain negative, significant, with a slightly larger magnitude. However, we still find no evidence that physical risk, as measured by our index, has a different impact on firm returns depending on the channel of exposure. Full results are available upon request.

	General		
	Model 1	Model 2	Model 3
Channel	0.09*	0.12	0.14
	(0.04)	(0.08)	(0.08)
Risk	−0.18***	−0.16***	−0.17***
	(0.04)	(0.04)	(0.04)
Channel x Risk	−0.00		−0.00
	(0.00)		(0.00)
Exposure		−0.05	−0.04
		(0.04)	(0.04)
Channel x Exposure		−0.01	−0.01
		(0.01)	(0.01)
Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Industry x Time FE	Yes	Yes	Yes
Num. obs.	136301	136301	136301
Num. ISIN	2826	2826	2826
R^2 (full model)	0.44	0.44	0.44
Adj. R^2 (full model)	0.29	0.29	0.29

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 9: Stock Returns and Aggregate Physical Risk Exposure: Estimation Results for Model with Channel Interaction 12

The table presents the estimation results of the model examining the relationship between financial returns and physical risk metrics—specifically, the hazard-level exposure and risk indices derived from earnings calls. In this setting, we include the interaction with the channel index to assess different response of stock returns depending on the fact that the hazard had direct or indirect effects to firms operations. The model is estimated at quarterly frequency and includes several balance sheet indicators as control variables. The analysis is conducted on a restricted sample of firms where stock returns, physical risk metrics, and balance sheet indicators are matched. Firm, time, and firm-time fixed effects are incorporated into the model, and standard errors are clustered at the firm level.

In conclusion, we find a significantly negative relationship between our measure of physical risk based on earning call transcripts and stock returns, driven by firms that face cold waves, floods, hurricanes and wind. This corroborates the hypothesis that physical risk exposure to different hazards is priced in equity markets and that investors might react to news of climate hazards leading to declines in stock returns. In contrast, there is no statistical evidence of an association between financial performance and the exposure channels.

6 Conclusions

This paper proposes a novel, scalable methodology to assess firms' exposure to physical climate risks by leveraging earnings call transcripts and state-of-the-art Natural Language Processing (NLP) techniques. We develop and deploy a multi-step framework, STORM, which integrates a taxonomy of climate hazards with lexical chaining, distributional semantics, and prompt-based classification using Large Language Models. We applied this approach to a large sample of over 3,000 firms worldwide, covering the period from 2001 to 2021. This enabled us to construct a novel dataset containing firm-level exposure metrics for four specific climate hazards—storms, cold waves, heatwaves, and wildfires—along with measures that differentiate between negative and positive impacts and classify exposure as direct or indirect.

Compared to the existing literature, our approach presents several key advancements. First, while most prior studies rely on asset-level geolocation data to estimate direct exposure to physical risk , we extract firm-level exposure metrics directly from unstructured textual data. This allows us to circumvent common data limitations, such as lack of facility-level information. Second, we go beyond traditional dictionary-based or machine learning sentiment classifiers by introducing a semantically enriched taxonomy combined with prompt-engineered generative models (GPT-3.5), called STORM, improving precision in classifying hazards, risks, and channels of impact . To our knowledge, this is the first framework to

integrate taxonomy development, lexical chaining, and prompt-based LLM reasoning to generate structured firm-level indicators from financial disclosures. Third, our methodology enables hazard-specific and sentiment-aware exposure analysis. Unlike previous approaches that often aggregate hazards or focus on a single type (mostly hurricanes or floods), we provide a multidimensional perspective and empirically demonstrate its relevance in explaining financial outcomes. Notably, we show that climate risk is predominantly transmitted through indirect channels, such as supply chain disruptions or market demand shifts.

Our findings have significant policy implications. The first key policy implication of this study is that all climate hazards considered in our analysis are financially material, as they have a measurable impact on stock returns. Importantly, we find that an increase in firms' risk index negatively affects valuations, with this effect remaining consistent across all hazard types. Adding to this complexity, firms' exposure is not confined to a single hazard; rather, they are typically affected by multiple risks simultaneously. This suggests that firms may need comprehensive and adaptive strategies to effectively manage these threats. These findings reinforce the urgency for regulators and financial actors to integrate climate risk into investment strategies and stress-testing frameworks to better anticipate financial market responses to extreme weather events.

Secondly, we show the predominant transmission channel of physical climate risk is indirect, with climate shocks impacting firms through disruptions in supply chains, shifts in market demand, or disturbances in broader economic networks rather than direct damage to physical assets. This has crucial implications for financial stability assessments conducted by central banks, as traditional approaches focusing solely on firms' physical assets may severely underestimate systemic risk. A comprehensive evaluation of firms' exposure must extend beyond their production processes to encompass vulnerabilities within their supply chains and end markets. Without detailed and reliable data on these interconnections, regulators may struggle to accurately assess the cascading effects of climate-related disruptions across industries and financial markets. Therefore, enhancing data availability on firms' commercial networks, supplier dependencies, and customer concentrations should be a priority for both policymakers and financial institutions. For the private sector, particularly in credit risk and market risk assessments, our findings highlight the need for more sophisticated modelling of physical risk exposure. Financial institutions assessing corporate borrowers or investment portfolios should assess exposure through supplier networks and customer bases into their risk frameworks, as indirect exposure could significantly impact firms' financial performance. Improved disclosure requirements and standardized reporting on climate-related dependencies would help mitigate the uncertainty surrounding indirect risks, benefiting investors, lenders, and insurers alike.

Finally, while we focus here on earnings call transcripts, the STORM framework can also be applied to extract climate-related information from sources beyond earnings call transcripts, such as non-financial reports (Hub, 2024; A.-I. Moreno and Caminero, 2022), 10-K filings (Berkman, Jona, and Soderstrom, 2024; Koelbel et al., 2020) and Pillar 3 disclosures (Á. I. Moreno and Caminero, 2022). These alternative sources provide valuable insights into firms' climate risk exposure and financial materiality, complementing earnings call analysis. This can be effectively implemented by developing a tailored taxonomy specific to each reporting framework under analysis, ensuring accurate extraction of relevant climate-related insights. This approach is particularly valuable in the context of a growing regulatory emphasis[9] on corporate environmental disclosure, which is expected to significantly expand the availability of standardized climate-related data.

7 Appendix

7.1 Extended Boolean Models

A fundamental approach in Information Retrieval (IR) is the linear scan across documents to identify matches, allowing the use of regular expressions. This approach is efficient when the collection of documents is not large, and the search criteria are simple and well-defined (Manning, 2009). The standard Boolean model can be enhanced by incorporating pattern matching and proximity-based approaches, leading to what we refer to as an Extended Boolean Model. Pattern matching allows the identification of sequences of characters or patterns that meet specified criteria, beyond the simple presence or absence

[9]Legislative initiatives such as the EU Corporate Sustainability Reporting Directive (CSRD) and the SEC's proposed climate disclosure rules aim to enhance transparency by requiring firms to disclose climate risks and their financial implications.

of terms. The proximity approach, on the other hand, focuses on the closeness of terms within documents (Manning, 2009), improving the retrieval of documents where search terms appear close together, indicating a stronger relationship and higher relevance to the query. This method is useful for searches where the context in which terms appear together is crucial for identifying relevant documents.

For SKM, where we want to find a word w and its plural form, the regular expression can be defined as follows:

$$\text{regex}_{\text{SKM}} = \texttt{r'\textbackslash b'} + w + \texttt{r's?\textbackslash b'}$$

This regular expression matches the exact word w or its plural form (indicated by $s?$), ensuring that the match occurs at word boundaries ($\backslash b$) to avoid partial matches within longer words.

NHM involves finding documents where two words, w and word_tgt, appear within a certain distance from each other. The regular expression for this type of search is more complex:

$$\begin{aligned}\text{regex}_{\text{NHM}} = \texttt{r'\textbackslash b(?:'} + w + \texttt{r's?\textbackslash W+(?:\textbackslash w+\textbackslash W+)\{0,40\}?(?:'} \\ + \text{word_tgt} + \texttt{r')|(?:'} + \text{word_tgt} + \texttt{r')\textbackslash W+(?:\textbackslash w+\textbackslash W+)\{0,40\}?'} \\ + w + \texttt{r's?)\textbackslash b'}\end{aligned} \tag{13}$$

This expression allows for matching documents where w and word_tgt are within 40 words of each other, considering both orders of appearance and allowing for the plural form of w.

7.2 Skip-Gram model of Word2vec

Word2Vec is a neural network-based approach for learning vector representations of words, known as word embeddings. This model captures syntactic and semantic word relationships. It has two architectures, Continuous Bag of Words (CBOW) and Skip-Gram. CBOW generates its output by averaging the context of surrounding words, while Skip-Gram constructs unique semantic representations for each word. This means that when a word is used in two sentences with differing meanings, the CBOW model merges these contexts into an average representation, failing to recognize the distinct meanings. In contrast, the Skip-Gram model identifies the word as unique in each instance, creating individual semantic representations for every context (Choudhary and Beniwal, 2021). In the context of identifying a taxonomy for climate events, the Skip-Gram process can be advantageous since precise differentiation in word usage is crucial. The Skip-Gram model of Word2Vec aims to predict context words for a given target word, maximizing the likelihood of context words within a certain window. The objective function for the Skip-Gram model is given by:

$$\mathcal{S} = \frac{1}{B}\sum_{b=1}^{B}\sum_{-D\le j\le D, j\neq 0} \log p(w_{b+j}|w_b) \tag{14}$$

where B is the total number of words in the corpus, D is the size of the context window, and $p(w_{b+j}|w_b)$ is the probability of observing a context word w_{b+j} given the current word w_b.

The probability $p(w_O|w_I)$, of observing the context word w_O given the input word w_I, is computed using the softmax function:

$$p(w_O|w_I) = \frac{\exp({v'_{w_O}}^T v_{w_I})}{\sum_{w=1}^{W}\exp({v'_w}^T v_{w_I})} \tag{15}$$

where v_{w_I} and v'_{w_O} are the input and output vector representations of the words, and W is the vocabulary size. The Skip-Gram model variant optimizes the average log probability of context words given a target word, using the softmax function to compute probabilities.

After training, each word in the vocabulary is represented as a vector in a high-dimensional space. To measure the semantic similarity between two words, we employ cosine similarity as a metric for similarity between two terms. It is formalized as follows

$$CosineSimilarity = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} \tag{16}$$

where $\mathbf{A}$ and $\mathbf{B}$ represent the two vectors between which we want to calculate the cosine similarity, $\cdot$ denotes the dot product between the vectors. $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ are the Euclidean norms (magnitudes) of the

vectors **A** and **B** respectively. This formula computes the cosine of the angle between the two vectors, which indicates their similarity in direction, irrespective of their magnitudes. In the context of Word2Vec, cosine similarity between two-word vectors indicates how similar or related the corresponding words are in meaning. It is a commonly used metric for comparing word embeddings and identifying similar words within a given corpus.

7.3 Prompt Engineering

Prompts are bodies of text used as the textual input to the generative AI model to generate a response. Effective prompt design, or "prompting", coaxes the model into performing analyses typically outside its intended output scope, such as sentiment analysis, pattern recognition, or text categorization. Large Language Models (LLMs) and chatbots generate text in response to input by calculating token generation probabilities and selecting the optimal continuation. However, they are not inherently designed for analytical tasks. By embedding instructions within the prompts and requesting a specific standardized output structure, we guide the model's output, facilitating the extraction of structured data from unstructured text and conducting sophisticated text analyses with elevated performance levels.

As described in Section 4.2 for our study, we devise three zero-shot prompts, each with explicit instructions detailing: a) the tasks to be executed, including descriptions of the indicators to be generated, b) the format of the desired output, which is a JSON document with labeled variables, c) the analytical context, encompassing a brief definition of an earnings conference call, the paragraphs subject to analysis, the keyword identifying the hazard type, accompanied by a succinct hazard group description and the NACE 1 prevalent economic sector of the company.

Full prompts are available in the following tables:

Role	Content
System	"You are a virtual assistant expert in extracting information from a given input text document."
User	"Your task is to extract information from a text provided as input by the user. - The user will ask you a question, provide you with response instructions, and with the text to analyze. - The response instructions will be provided through a numbered list. - The text will be between triple quotes and will be in English. - The text will be a transcription of a passage from a video call organized by a listed company to discuss its financial results with financial analysts and shareholders. - The text may report part of a speech by company representatives, a question from other participants in the meeting, or an exchange of questions and answers among several people. When responding to the questions, please consider that the company is operating in the following sector: {NACE}."
Assistant	"Perfect. Please provide your question, response instructions, and text."
User	"Question: In the text reported between triple quotes, does it talk about or mention {KEYWORD}, a weather/atmospheric and/or climate event? Reason step-by-step to make sure you provide a correct answer. Below are the response instructions: 1) Your response has to always include a JSON object, 2) The key of the JSON object is 'bool', 3) For the 'bool' key: If mentions of weather and/or climate events have been found, respond in this way: 'bool': True. Otherwise, if none of these events are mentioned in the text, respond: 'bool': False. Respond False even when the text includes the word {KEYWORD}, but the speaker does not refer to a weather/atmospheric/climate event (for instance the word is used in a metaphorical sense/colloquialism or it is a name of a person, location, product or project). Below is the text to be analyzed: {TEXT}"

Table 10: GPT prompt - Validation of Climate Hazard Exposure

The table reports the prompt developed to instruct the LLM for completing Task 1: the validation of hazard exposure. The prompt mocks an initial interaction between the Chatbot and a human User, where the User provides a question and the necessary instructions to respond correctly. The main roles are "System" (general instructions for the role the LLM has to fulfill), "User" (the agent interacting with the AI assistant), and "Assistant" (the AI assistant).

Role	Content
System	"You are a virtual assistant expert in extracting information from a given input text document."
User	"Your task is to extract information from a text provided as input by the user. - The user will ask you a question, provide you with response instructions, and with the text to analyze. - The response instructions will be provided through a numbered list. - The text will be between triple quotes and will be in English. - The text will be a transcription of a passage from a video call organized by a listed company to discuss its financial results with financial analysts and shareholders. - The text may report part of a speech by company representatives, a question from other participants in the meeting, or an exchange of questions and answers among several people. When responding to the questions, please consider that the company is operating in the following sector: {NACE}."
Assistant	"Perfect. Please provide your question, response instructions, and text."
User	"Question: The text reported between triple quotes mentions {KEYWORD}, a climate/weather/atmospheric event. Does this type of event represent a risk or an opportunity for the company and its business activities? Or are these events irrelevant to the company's activities? Reason step-by-step to ensure that you infer a correct answer from the text. Below are the response instructions: 1) Your response has to always include a JSON object, 2) The key of the JSON object is 'sentiment', 3) for the key 'sentiment': if these events are relevant for the company and they represent an opportunity for its business, respond in this way: 'sentiment': positive. If the events are relevant for the company and they represent a risk for its business, respond: 'sentiment': negative. If it is clearly stated that the events have no impacts of any kind on the company's activities or if there is not enough information, respond: 'sentiment': zero. Below is the text to be analyzed: {TEXT}"

Table 11: GPT prompt - Risk versus Opportunities

The table reports the prompt developed to instruct the LLM for completing Task 2: the classification of risks and opportunities derived from hazard exposure. The prompt mocks an initial interaction between the Chatbot and a human User, where the User provides a question and the necessary instructions to respond correctly. The main roles are "System" (general instructions for the role the LLM has to fulfill), "User" (the agent interacting with the AI assistant), and "Assistant" (the AI assistant).

Role	Content
System	"You are a virtual assistant expert in extracting information from a given input text document."
User	"Your task is to extract information from a text provided as input by the user. - The user will ask you a question, provide you with response instructions, and with the text to analyze. - The response instructions will be provided through a numbered list. - The text will be between triple quotes and will be in English. - The text will be a transcription of a passage from a video call organized by a listed company to discuss its financial results with financial analysts and shareholders. - The text may report part of a speech by company representatives, a question from other participants in the meeting, or an exchange of questions and answers among several people. When responding to the questions, please consider that the company is operating in the following sector: {NACE}."
Assistant	"Perfect. Please provide your question, response instructions, and text."
User	"Question: The text reported between triple quotes mentions {KEYWORD}, a climate/weather/atmospheric event. How does this event impact the company's activities? Categorize the initial channel through which the events impact the company. Here are the available response options: 1) indirect impacts: the events do not directly impact the company's assets but cause shocks in its supply chain or end markets. This channel may include shocks in the demand for goods and services sold by the company or shocks in the supply of goods and services purchased by the firm to conduct its activities. 2) impacts on production: the events directly hit the company in its physical locations, affecting its production, assets and/or organizational/business activities. Examples may include direct damage to its assets, additional costs and/or delays in production/activities. 3) there is not enough information to determine the channel through which the events impact the company. Reason step-by-step to ensure that you provide a correct answer. Keep in mind that the examples provided are indicative and not exhaustive. Find below the response instructions: 1) Your response has to always include a JSON object, 2) The key of the JSON object is 'channel', 3) for the key 'channel': If mentions have been found regarding the impact of the event through the indirect channel (option 1), respond in this way: 'channel': indirect. If instead the text mentions a direct impact of the event on the company (option 2), respond: 'channel': direct. If there is not enough information to infer the channel through which the event affects the company, or if the events have no impact on the company's activities, or if it is not possible to identify the events in the text (option 3), respond: 'channel': NULL. Below is the text to be analyzed: {TEXT}"

Table 12: GPT prompt - Channels of Corporate Exposure

The table reports the prompt developed to instruct the LLM for completing Task 3: the classification of exposure channels. The prompt mocks an initial interaction between the Chatbot and a human User, where the User provides a question and the necessary instructions to respond correctly. The main roles are "System" (general instructions for the role the LLM has to fulfill), "User" (the agent interacting with the AI assistant), and "Assistant" (the AI assistant).

7.4 Transcript and firm statistics

Country	Firms Number	Transcripts Number	Percentage of Tot. Transcripts	Covered NACE 1	Average Transcripts Length	Average Dictionary Length	Average Paragraphs Number	Std. Dev. Paragraph Number	Average Paragraphs Length	Std. Dev. Paragraph Length
	10^0	10^0	%	10^0	10^3	10^3	10^0	10^0	10^0	10^0
Australia	20	327	0.3	10	67.7	1.8	113	39.4	428	7.6
Austria	9	298	0.3	4	57.0	1.5	99	42.2	427	7.4
Belgium	13	383	0.4	4	67.8	1.8	116	43.3	426	7.6
Bermuda	3	136	0.1	3	58.7	1.6	98	35.8	428	7.6
Brazil	16	888	0.9	5	50.3	1.4	87	33.4	426	7.4
Canada	269	6896	7.2	16	57.7	1.5	91	33.4	434	7.6
Cayman Islands	2	119	0.1	2	53.0	1.5	88	36.0	431	7.6
Chile	4	149	0.2	3	43.3	1.3	76	22.9	424	7.4
China	18	349	0.4	7	46.2	1.3	80	27.4	427	7.6
Colombia	2	41	0.0	1	55.1	1.6	96	25.4	429	7.3
Cyprus	1	11	0.0	1	45.3	1.5	80	11.7	426	7.9
Czech Republic	3	135	0.1	2	53.7	1.4	92	37.9	428	7.6
Denmark	21	809	0.8	6	61.3	1.6	105	36.5	428	7.5
Faroe Islands	1	18	0.0	1	35.8	1.0	67	12.7	429	7.5
Finland	24	838	0.9	8	53.8	1.4	89	33.4	429	7.5
France	65	1602	1.7	14	70.8	1.9	128	48.2	422	7.5
Germany	68	2583	2.7	11	68.2	1.8	118	45.6	426	7.6
Gibraltar	1	34	0.0	1	53.7	1.6	101	48.5	422	7.5
Greece	10	278	0.3	4	52.9	1.5	88	30.1	426	7.6
Hong Kong	10	200	0.2	7	40.1	1.2	72	34.4	424	7.6
Hungary	1	56	0.1	1	51.2	1.5	86	24.3	426	7.5
Iceland	2	18	0.0	1	32.0	0.9	52	14.1	435	7.2

Continued

Table 13: Firms and Transcripts by Country

The table shows descriptive statistics for the transcript database by firm headquarters' country. Data by country includes (in column order) the number of firms, the number of available transcripts, the percentage of transcripts over the total sample, the number of different NACE 1 economic sectors across firms, the average length of transcripts in characters, the average length of unique words included in a transcript (dictionary), the mean and standard deviation of transcripts' length in paragraphs, the mean and standard deviation of paragraphs' length in characters.

Table 14: Firms and Transcripts by Country

Continued

Country	Firms Number	Transcripts Number	Percentage of Tot. Transcripts	Covered NACE 1	Average Transcripts Length	Average Dictionary Length	Average Paragraphs Number	Std. Dev. Paragraph Number	Average Paragraphs Length	Std. Dev. Paragraph Length
	10^0	10^0	%	10^0	10^3	10^3	10^0	10^0	10^0	10^0
India	10	305	0.3	4	83.5	1.9	128	34.0	437	7.5
Ireland	22	818	0.8	8	69.2	1.8	118	35.8	427	7.7
Israel	6	230	0.2	3	62.4	1.6	99	31.7	434	7.7
Italy	36	954	1.0	7	58.1	1.6	101	45.8	426	7.5
Japan	19	1427	1.5	7	34.1	0.9	51	46.9	456	7.6
Luxembourg	15	383	0.4	6	57.1	1.5	97	33.9	428	7.6
Malta	3	42	0.0	1	50.1	1.4	85	32.0	429	7.4
Mexico	3	91	0.1	2	57.6	1.5	96	20.3	427	7.5
Missing	143	5622	5.8	5	59.0	1.6	98	38.2	430	7.7
Netherlands	35	1278	1.3	12	70.1	1.8	119	45.0	427	7.7
New Zealand	1	9	0.0	1	84.4	2.4	143	19.0	431	7.5
Norway	40	1004	1.0	9	41.6	1.2	73	34.6	427	7.3
Philippines	2	27	0.0	2	66.9	1.8	110	28.0	429	7.8
Poland	9	104	0.1	5	45.3	1.3	79	23.7	427	7.3
Portugal	5	190	0.2	5	56.7	1.6	103	46.4	422	7.5
Russia	19	434	0.5	10	54.1	1.5	90	35.7	431	7.5
S. Korea	7	553	0.6	2	54.4	1.4	89	34.1	430	7.5
South Africa	5	105	0.1	2	76.2	2.1	140	46.2	420	7.6
Spain	29	890	0.9	10	57.0	1.6	101	41.8	425	7.5
Sweden	64	1841	1.9	14	58.3	1.4	97	32.2	430	7.5
Switzerland	37	1001	1.0	5	68.2	1.8	118	41.1	426	7.6
Taiwan	4	222	0.2	2	53.2	1.3	82	26.4	435	7.5
Thailand	1	25	0.0	1	33.0	1.1	55	15.6	436	7.3
UAE	2	49	0.1	2	31.5	1.1	56	24.3	425	7.7
USA	1886	57960	60.2	20	59.0	1.6	95	33.6	432	7.7
Ukraine	1	18	0.0	1	57.0	1.3	83	24.7	447	7.2
United Kingdom	176	4570	4.7	16	65.0	1.8	117	48.8	424	7.6
Virgin Islands	1	12	0.0	1	48.2	1.5	81	15.8	431	7.7
TOT	3144	96332	100.0	20	59.1	1.6	97	37.8	431	7.7

NACE 1	Firms Number	Transcripts Number	Percentage of Tot. Transcripts	Covered Countries	Average Transcripts Length	Average Dictionary Length	Average Paragraphs Number	Std. Dev. Paragraph Number	Average Paragraphs Length	Std. Dev. Paragraph Length
	10^0	10^0	%	10^0	10^3	10^3	10^0	10^0	10^0	10^0
A_AGRI	18	447	0.5	9	55.4	1.5	89	29.1	432	7.7
B_MINING	216	6402	6.6	18	58.9	1.5	95	38.5	432	7.7
C_MANUF.	1123	35694	37.1	36	60.6	1.6	99	38.3	431	7.7
D_ELEC.	96	3196	3.3	26	56.8	1.5	94	41.5	430	7.5
E_WATER	26	723	0.8	8	61.8	1.7	104	40.9	429	7.7
F_CONSTR.	75	1994	2.1	16	59.2	1.5	97	34.8	431	7.7
G_TRADE	229	7048	7.3	24	60.4	1.6	98	39.0	432	7.7
H_TRANSP	120	4083	4.2	22	59.9	1.6	98	41.2	431	7.6
I_HOTEL	61	1716	1.8	10	54.4	1.5	90	33.3	430	7.6
J_ICT	471	13785	14.3	33	58.4	1.6	97	36.8	430	7.7
K_FIN	15	443	0.5	5	60.0	1.6	98	43.3	430	7.6
L_RE	123	3674	3.8	13	60.4	1.6	98	34.4	431	7.6
M_PROF	211	5031	5.2	14	53.4	1.5	89	35.4	430	7.7
N_ADMIN	71	1952	2.0	14	58.4	1.5	97	36.3	430	7.6
O_P. ADM	1	29	0.0	1	43.3	1.3	75	16.3	427	7.7
P_EDU	15	548	0.6	3	59.5	1.6	97	32.4	430	7.7
Q_HUMAN	59	1741	1.8	7	59.9	1.6	97	34.6	431	7.7
R_ARTS,	30	992	1.0	7	49.7	1.4	81	34.2	432	7.6
S_OTH. SERV	13	358	0.4	4	62.3	1.6	101	29.4	431	7.7
NA	171	6476	6.7	5	58.8	1.6	97	37.5	430	7.7
TOT	3144	96332	100.0	50	59.1	1.6	97	37.8	431	7.7

Table 15: Firms and Transcripts by NACE 1 economic sector

The table shows descriptive statistics for the transcript database by firms' NACE 1 economic sector. Data by sector includes (in column order) the number of firms, the number of available transcripts, the percentage of transcripts over the total sample, the number of different headquarters' countries across firms, the average length of transcripts in characters, the average length of unique words included in a transcript (dictionary), the mean and standard deviation of transcripts' length in paragraphs, the mean and standard deviation of paragraphs' length in characters.

Year	Firms Number	Transcripts Number	Percentage of Tot. Transcripts	Covered NACE 1	Covered Countries	Average Transcripts Length	Average Dictionary Length	Average Paragraphs Number	Std. Dev. Paragraph Number	Average Paragraphs Length	Std. Dev. Paragraph Length
	10^0	10^0	%	10^0	10^0	10^3	10^3	10^0	10^0	10^0	10^0
2001	48	72	0.1	9	8	70.8	1.7	110	36.6	440	7.6
2002	636	1693	1.8	19	25	64.1	1.6	95	37.5	451	7.6
2003	930	3388	3.5	19	34	63.8	1.6	97	39.5	438	7.6
2004	1030	4108	4.3	19	34	64.3	1.6	99	36.5	436	7.6
2005	1198	4492	4.7	19	36	64.3	1.6	100	37.2	434	7.6
2006	1285	4764	4.9	19	38	62.8	1.6	98	37.8	433	7.7
2007	1380	5022	5.2	19	40	62.1	1.6	97	39.2	434	7.6
2008	1461	5383	5.6	19	39	61.9	1.6	98	38.9	432	7.7
2009	1414	5287	5.5	19	40	60.0	1.5	96	38.9	431	7.7
2010	1436	5316	5.5	19	41	60.1	1.5	97	37.9	431	7.7
2011	1530	5520	5.7	19	40	59.5	1.6	97	38.3	430	7.7
2012	1513	5286	5.5	19	40	59.1	1.6	97	38.0	429	7.7
2013	1305	4641	4.8	19	39	59.7	1.6	99	37.4	429	7.7
2014	1390	4838	5.0	20	39	58.7	1.6	102	38.3	425	7.8
2015	1358	4604	4.8	20	38	58.4	1.6	101	37.3	424	7.8
2016	1273	4262	4.4	20	41	58.0	1.6	101	38.0	423	7.8
2017	1545	5056	5.2	20	48	55.1	1.5	94	37.7	431	7.6
2018	1599	5686	5.9	20	49	54.1	1.5	92	35.6	432	7.5
2019	1591	5678	5.9	20	50	53.7	1.5	91	36.1	431	7.5
2020	1563	5722	5.9	20	49	55.9	1.6	95	37.4	431	7.6
2021	1491	5514	5.7	20	49	55.1	1.6	95	36.4	428	7.6
TOT	3144	96332	100.0	20	50	59.1	1.6	97	37.8	431	7.7

Table 16: Firms and Transcripts by year

The table shows descriptive statistics for the transcript database by year. Data by year includes (in column order) the number of firms, the number of available transcripts, the percentage of transcripts over the total sample, the number of different NACE 1 economic sectors and of different headquarters' countries across firms, the average length of transcripts in characters, the average length of unique words included in a transcript (dictionary), the mean and standard deviation of transcripts' length in paragraphs, the mean and standard deviation of paragraphs' length in characters.

Statistic	N	Mean	St. Dev.	Min	Max
Exposure	136,301	0.1	1.3	−0.2	21.3
Risk	136,301	0.04	1.3	−50.9	45.1
Channel	136,301	−0.03	1.3	−63.0	50.5
Returns(%)	136,301	2.0	21.1	−57.1	84.1
lnAssets	136,301	13.6	2.3	0.0	20.5
Debt/Assets	136,301	0.3	0.2	0.0	1.5
Cash/Assets	136,301	0.2	0.2	0.0	0.9
CapEx/Assets	136,301	0.1	0.1	0.0	0.4
EBIT/Assets	136,301	−0.02	0.3	−2.1	0.4
PPE/Assets	136,301	0.3	0.3	0.0	1.0
R&D/Assets	136,301	0.1	0.1	0.0	1.1
SalesGrowth	136,301	16.5	57.5	−73.0	453.2

Table 17: Summary statistics

The table provides summary statistics for the variables included in the restricted firm-level dataset, where stock returns, balance sheet indicators, and physical risk metrics are matched at quarterly frequency.

Percentiles of Physical risk Scores								Correlation Matrix of Physical risk Scores											
	1th	5th	25th	50th	75th	95th	99th		Exposure				Risk/Opp				Channel			
Exposure									Cold Waves	Floods	Heat wave	Wildfires	Cold Waves	Floods	Heat wave	Wildfires	Cold Waves	Floods	Heat wave	Wildfires
Cold wave	-0.105	-0.105	-0.105	-0.105	0.003	0.439	1.309	Cold wave	1	.	.	.	.	.	.	.	.	.	.	.
Floods	-0.152	-0.152	-0.152	-0.152	-0.001	0.601	1.429	Floods	.40***	1	.	.	.	.	.	.	.	.	.	.
Heat wave	-0.082	-0.082	-0.082	-0.082	-0.082	0.338	1.456	Heat wave	.30***	.30***	1	.	.	.	.	.	.	.	.	.
Wildfires	-0.040	-0.040	-0.040	-0.040	-0.040	0.246	0.820	Wildfires	.18***	.18***	.17***	1	.	.	.	.	.	.	.	.
Risk/Opp									Cold Waves	Floods	Heat wave	Wildfires	Cold Waves	Floods	Heat wave	Wildfires	Cold Waves	Floods	Heat wave	Wildfires
Cold wave	-0.541	-0.093	-0.004	-0.004	-0.004	0.176	0.444	Cold wave	-.40***	.01	-.02	-.04	1	.	.	.	.	.	.	.
Floods	-0.203	-0.103	-0.064	-0.064	-0.024	0.335	0.893	Floods	.24***	.55***	.17***	.09***	.09***	1	.	.	.	.	.	.
Heat wave	-0.226	-0.122	-0.019	-0.019	-0.019	0.085	0.705	Heat wave	.11***	.15***	.41***	.04	.01	.04	1	.	.	.	.	.
Wildfires	-0.273	-0.017	-0.017	-0.017	-0.017	-0.017	0.495	Wildfires	.12***	.09***	.11***	.66***	.05	.10***	.02	1	.	.	.	.
Channel									Cold Waves	Floods	Heat wave	Wildfires	Cold Waves	Floods	Heat wave	Wildfires	Cold Waves	Floods	Heat wave	Wildfires
Cold wave	-0.724	-0.188	0.027	0.027	0.027	0.134	0.349	Cold wave	-.61***	-.12***	-.09***	-.09***	.85***	.00	-.05	.02	1	.	.	.
Floods	-0.641	-0.227	0.028	0.028	0.028	0.091	0.410	Floods	-.18***	-.35***	-.12***	-.08***	.01	.15***	-.16***	.01	.12***	1	.	.
Heat wave	-0.830	-0.113	0.030	0.030	0.030	0.030	0.173	Heat wave	-.24***	-.25***	-.65***	-.04	.01	-.10***	-.31***	-.01	.09***	.20***	1	.
Wildfires	-0.406	0.004	0.004	0.004	0.004	0.004	0.414	Wildfires	-.01	-.02	-.01	-.09***	.02	-.01	-.02	.25***	.06*	.08***	.01	1

Table 18: Descriptive statistics of Physical risk indicators

The table presents key summary statistics for the physical risk indicators, averaged over time for each firm. On the left, it displays the percentiles of the distribution for each metric, while on the right, it includes the correlation matrix. Hazard labels have been shortened for clarity and readability.

7.5 Comparison of GPT's performance with other models

Methodology	Class	Precision	Recall	F1	Balanced Accuracy
GPT Model	Neutral	0.54	0.55	0.54	0.71
	Risk	0.86	0.78	0.82	0.82
	Opportunity	0.71	0.85	0.77	0.87
Distilled LLM	Neutral	0.23	0.31	0.26	0.52
	Risk	0.66	0.65	0.66	0.62
	Opportunity	0.50	0.38	0.43	0.63
Dictionary-based over Paragraphs	Neutral	0.23	0.20	0.21	0.51
	Class: Risk	0.65	0.64	0.65	0.61
	Opportunity	0.39	0.44	0.41	0.61
Dictionary-based over document	Neutral	0.27	0.01	0.03	0.50
	Risk	0.57	0.77	0.66	0.53
	Opportunity	0.27	0.27	0.27	0.52

Table 19: Models' performance in the risk vs opportunity indicator

Table 19 shows results of a comparison exercise across three different model applications in measuring risks and opportunities. Performances are measured as in 3 over a sample of manually annotated paragraphs. The included models are OpenAi's GPT 3.5 Turbo - the model we deploy in our analyses; 2) a distilled ClimateBERT model from (Bingler et al., 2022); 3) a classic dictionary-based sentiment score from Loughran and McDonald (2011), measured both at paragraphs and at document levels.

The GPT model outperforms both methodologies substantially in all classes and performance indicators. While we expect a significant difference in performances when comparing generative large language models with dictionary-based approaches, the comparison with ClimateBert might seem surprising. In reality, the ClimateBERT performance might resent the different nature of its training data (firm's annual reports) when compared to our transcript database.

As discussed in Section 2.3, to properly deploying a distilled model researchers have to take into account additional costs in terms of parameter training over relevant data sources. On the other hand, larger models such as GPT offer a plug-and-play solution, at the expenses of a slightly lower performance and higher running costs. On this topic, the research frontier seems to be moving decidedly toward distilled models, also given the increasingly significant environmental costs that LLMs' deployments and maintenance cause.

7.6 Physical Risk Indexes and Firm Characteristics

In this section we examine the relationship between our two physical risk measures - the aggregate exposure index and the risk index - and various firm characteristics. Next, we conduct a variance decomposition analysis to quantify how much of the variation in the physical risk indexes are explained by different sets of fixed effects and firm-level characteristics.

Table 20 presents the result of the regression relating the two physical risk indices derived from earnings calls and firm characteristics. The coefficient of total assets is positive and statistically significant across models, suggesting that larger firms tend to face higher exposure to physical hazards and greater risk, possibly because they operate more extensive infrastructure and have broader geographical footprints. Similarly, Property, Plant, and Equipment (PPE) show a strong positive correlation with both exposure and risk, implying that firms with a greater proportion of tangible assets are more vulnerable to physical risks, likely due to their dependence on infrastructure-sensitive activities. Conversely, some financial characteristics appear to have weaker relationships with exposure and risk. For example, the debt-to-assets ratio and cash holdings show no strong correlation with physical risk indicators. These results are consistent with the findings of Q. Li et al. (2020).

Table 21 breaks down the analysis by hazard. The results show that firm size and PPE remain important predictors of exposure across different hazard categories however with some heterogeneity. In particular, total assets is an important driver of exposure across all hazard types, but its relationship

	Model 1	Model 2
	Exposure	Risk
logASSETS	0.04***	0.02***
	(0.00)	(0.00)
DEBT/ASSETS	0.03	0.02
	(0.04)	(0.03)
CASH/ASSETS	−0.07	−0.05*
	(0.04)	(0.02)
CAPEX/ASSETS	−0.32	0.04
	(0.17)	(0.12)
EBIT/ASSETS	−0.00	−0.01
	(0.02)	(0.01)
PPE/ASSETS	0.35***	0.15***
	(0.06)	(0.04)
R&D/ASSETS	0.11*	0.07*
	(0.05)	(0.03)
SALES growth	−0.00*	−0.00***
	(0.00)	(0.00)
Time FE	Yes	Yes
NACE-2 x Time FE	Yes	Yes
Num. obs.	136301	136301
Num. TIME	84	84
Num. NACE-2 x TIME	6637	6637
R^2 (full model)	0.13	0.10
Adj. R^2 (full model)	0.08	0.05

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 20: Firm-level determinants of Exposure and Risk Index with firm characteristics

The table presents the estimation results of the regression relating the aggregate exposure and risk indices derived from earnings calls and firm characteristics. The model is estimated at quarterly frequency. The analysis is conducted on a restricted sample of firms where stock returns, physical risk metrics, and balance sheet indicators are matched. Time and NACE-2-time fixed effects are incorporated into the model, and standard errors are clustered at the firm level.

with risk is only significant for heatwaves & droughts, flood, hurricanes & wind, and wildfires to a lesser extent. Similarly, PPE is more strongly associated with negative exposure to heatwaves & droughts and wildfires.

The variance decomposition analysis follows the approach of Sautner et al. (2023), and the results are presented in Table 22. The findings indicate that industry × time fixed effects account for a substantial share of the model's explanatory power (R^2), suggesting that industry-wide shocks and time-varying sectoral characteristics play a critical role in shaping firms' exposure to physical risk. Firm-level characteristics also contribute significantly to the explained variance, influencing both the Exposure Index and Risk Index. Overall, the results show that both firm-specific and industry-level factors are primary drivers of corporate exposure to climate hazards, collectively explaining between 30% and 50% of the model's R^2, depending on the dependent variable analyzed. These findings diverge from existing literature. While Q. Li et al. (2020) and Sautner et al. (2023) attribute most of the variation in their physical risk measures to firm-level characteristics, our results highlight a greater role for industry-wide factors, emphasizing the importance of sectoral exposure in shaping corporate vulnerability to climate hazards.

	cold wave		heat wave&drought		flood,hurricanes&wind		wildfire	
	Exposure	Risk	Exposure	Risk	Exposure	Risk	Exposure	Risk
log ASSETS	0.02***	−0.00	0.04***	0.02***	0.02***	0.01***	0.01**	0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
DEBT/ASSETS	0.01	−0.03	0.07*	0.04*	−0.05	0.00	−0.02	−0.06***
	(0.03)	(0.05)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
CASH/ASSETS	−0.07**	−0.02	−0.02	−0.04	−0.04	−0.01	−0.03	−0.03
	(0.03)	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
CAPEX/ASSETS	−0.22	0.20	−0.18	−0.02	−0.22	−0.04	−0.05	−0.08
	(0.15)	(0.14)	(0.16)	(0.12)	(0.19)	(0.09)	(0.12)	(0.08)
EBIT/ASSETS	−0.02	−0.01	0.03	0.01	−0.01	−0.02	−0.03	−0.04*
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
PPE/ASSETS	0.20***	0.03	0.22***	0.13***	0.23***	0.04	0.13*	0.12***
	(0.05)	(0.03)	(0.06)	(0.04)	(0.06)	(0.03)	(0.05)	(0.04)
R&D/ASSETS	0.04	0.01	0.10*	0.08*	0.05	0.01	0.01	−0.01
	(0.04)	(0.04)	(0.04)	(0.03)	(0.06)	(0.04)	(0.05)	(0.03)
SALES growth	0.00	−0.00	−0.00***	−0.00**	−0.00*	−0.00	0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NACE-2 x Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	136301	136301	136301	136301	136301	136301	136301	136301
Num. TIME	84	84	84	84	84	84	84	84
Num. NACE-2 x TIME	6637	6637	6637	6637	6637	6637	6637	6637
R^2 (full model)	0.10	0.04	0.12	0.11	0.10	0.09	0.07	0.07
Adj. R^2 (full model)	0.05	−0.01	0.08	0.06	0.05	0.05	0.02	0.02

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 21: Firm-level determinants of Exposure and Risk Index with firm characteristics by hazards

The table presents the estimation results of the regression relating the aggregate exposure and risk indices derived from earnings calls and firm characteristics by hazard. The model is estimated at quarterly frequency. The analysis is conducted on a restricted sample of firms where stock returns, physical risk metrics, and balance sheet indicators are matched. Time and NACE 2-time fixed effects are incorporated into the model, and standard errors are clustered at the firm level.

Model	Exposure R^2	Incremental R^2 (%)	Risk R^2	Incremental R^2 (%)
Full model	0.28	100.00	0.16	100.00
Only Time	0.02	7.64	0.01	7.63
Only NACE	0.04	14.13	0.01	5.20
Only NACE x Time	0.12	44.65	0.10	35.30
Unexplained (Firm-Level)	0.09	33.58	0.04	51.87

Table 22: Variance Decomposition of Exposure and Risk Index

The table presents the results of the variance decomposition analysis for the Exposure Index and Risk Index derived from earnings call transcripts. The analysis quantifies the contribution of different factors—firm fixed effects and firm covariates, time fixed effects, industry classification (NACE-2), and their interaction (NACE-2 × Time)—to the overall variation in firms' physical risk exposure. The first row reports the R^2 of the full model, which includes firm-level variables along with all fixed effects. Columns 1 and 3 display the R^2 of models with only one fixed effect at a time, isolating the explanatory power of each factor. The incremental R^2 (Columns 2 and 4) expresses the proportion of the full model's explanatory power attributed to each component. The unexplained firm-level variation is computed as the difference between the full model's R^2 and the sum of the individual fixed-effects model's R^2. The model is estimated at a quarterly frequency using a restricted sample of firms where stock returns, physical risk metrics, and balance sheet indicators are available. This analysis highlights the relative importance of firm-level, industry-specific, and time-driven factors in explaining corporate exposure to climate hazards.

Acknowledgements

We would like to thank Jakob Jones, Michele Tursi and Miles Tarrant for their support in the construction of the Earnings call database and contribution to the analysis. We also thank Michele Filannino for his guidance on GPT deployments and perceptive feedbacks; and Giovanna Bua andDaniel Kapp for their availability in sharing data and providing valuable insights. We also thank Leonardo Bortolan, Luca Zanin, Luca Taschini, Olimpia Carradori, Andrea Marcucci, Eugenia Frezza for their feedbacks on the paper. We also thank participants at the 2023 International Fintech Research Conference, the 2024 ITAL-IA Workshop, the 2024 Banque du France - Institute Bachelier GFRA conference and the Banca d'Italia – EABCN – EUI – CEPR joint conference on "The macroeconomic and financial dimensions of the green transition".

References

Addoum, Jawad M, David T Ng, and Ariel Ortiz-Bobea (2020). "Temperature shocks and establishment sales." In: *The Review of Financial Studies* 33.3, pp. 1331–1366.

Ardia, David et al. (2023). "Climate change concerns and the performance of green vs. brown stocks." In: *Management Science* 69.12, pp. 7607–7632.

Azure-AI-Services (2023). "OpenAI (gpt-35-turbo - version 2023-05-15) [Large language model]." In: URL: https://learn.microsoft.com/en-us/azure/ai-services/openai.

Barrot, Jean-Noël and Julien Sauvagnat (2016). "Input specificity and the propagation of idiosyncratic shocks in production networks." In: *The Quarterly Journal of Economics* 131.3, pp. 1543–1592.

Batten, S. (2018). "Climate Change and the Macro-Economy: A Critical Review." In: *Environmental Economics eJournal*. DOI: 10.2139/ssrn.3104554.

Battiston, S., Yannis Dafermos, and I. Monasterolo (2020). "Climate Risks and Financial Stability." In: *Monetary Economics: International Financial Flows*. DOI: 10.2139/ssrn.3748495.

BCBS (2021). *Climate-related financial risks–Measurement methodologies*.

Benincasa, Emanuela et al. (2024). "Rewiring Supply Chains Through Uncoordinated Climate Policy." In: *Available at SSRN 4994643*.

Berkman, Henk, Jonathan Jona, and Naomi S. Soderstrom (2024). "Firm-Specific Climate Risk and Market Valuation." In.

Bingler, Julia Anna et al. (2022). "Cheap talk and cherry-picking: What ClimateBert has to say on corporate climate risk disclosures." In: *Finance Research Letters* 47, p. 102776.

Bolton, Patrick and Marcin Kacperczyk (2021). "Do investors care about carbon risk?" In: *Journal of Financial Economics* 142.2, pp. 517–549. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2021.05.008. URL: https://www.sciencedirect.com/science/article/pii/S0304405X21001902.

Bortolan, Leonardo, Atreya Dey, and Luca Taschini (2024). *Volatile Temperatures and Their Effects on Equity Returns and Firm Performance*. Tech. rep. CESifo Working Paper.

Boubaker, Sabri et al. (2025). "Natural disasters, unnatural earnings: How do climate disasters impact earnings management?" In: *International Review of Financial Analysis*, p. 104043.

Bourdeau-Brien, Michael and Lawrence Kryzanowski (2017). "The impact of natural disasters on the stock returns and volatilities of local firms." In: *The Quarterly Review of Economics and Finance* 63, pp. 259–270.

Bressan, Giacomo et al. (2024). "Asset-level assessment of climate physical risk matters for adaptation finance." In: *Nature Communications* 15.1, p. 5371.

Bua, Giovanna et al. (2022). "Transition versus physical climate risk pricing in European financial markets: A text-based approach." In.

Byard, Donal, Mahmud Hossain, and Santanu Mitra (2007). "US oil companies' earnings management in response to hurricanes Katrina and Rita." In: *Journal of Accounting and Public Policy* 26.6, pp. 733–748.

Carleton, T. and S. Hsiang (2016). "Social and economic impacts of climate." In: *Science* 353. DOI: 10.1126/science.aad9837.

Choi, Darwin, Zhenyu Gao, and Wenxi Jiang (2020). "Attention to global warming." In: *The Review of Financial Studies* 33.3, pp. 1112–1145.

Choudhary, Kailash and Ruby Beniwal (2021). "Xplore Word Embedding Using CBOW Model and Skip-Gram Model." In: *2021 7th International Conference on Signal Processing and Communication (ICSC)*. IEEE, pp. 267–270.

Collier, Zachary A, Richard J Gruss, and Alan S Abrahams (2024). "How good are large language models at product risk assessment?" In: *Risk Analysis*.

"Datastream International" (2023). In: LSEG Workspace.

Donadelli, Michael et al. (2020). "Tornado activity, house prices, and stock returns." In: *The North American Journal of Economics and Finance* 52, p. 101162.

Engle, Robert F et al. (2020). "Hedging climate change news." In: *The Review of Financial Studies* 33.3, pp. 1184–1216.

Faccini, Renato, Rastin Matin, and George Skiadopoulos (2023). "Dissecting climate risks: Are they reflected in stock prices?" In: *Journal of Banking & Finance* 155, p. 106948.

FSB (2022). *Climate Scenario Analysis by Jurisdictions*.

Ghadge, Abhijeet, Hendrik Wurtmann, and S. Seuring (2019). "Managing climate change risks in global supply chains: a review and research agenda." In: *International Journal of Production Research* 58, pp. 44–64. DOI: 10.1080/00207543.2019.1629670.

"Greenhouse Gas Reporting Program (GHGRP)" (2023). In: Date accessed: [07 10, 2024]: U.S. Environmental Protection Agency Office of Atmospheric ProtectionU.S. Environmental Protection Agency Office of Atmospheric Protection. URL: www.epa.gov/ghgreporting.

Hassan, Tarek A et al. (2019). "Firm-level political risk: Measurement and effects." In: *The Quarterly Journal of Economics* 134.4, pp. 2135–2202.

— (2020a). *Firm-level exposure to epidemic diseases: COVID-19, SARS, and H1N1*. Working Paper 26971. National Bureau of Economic Research.

— (2020b). *The global impact of Brexit uncertainty*. Working Paper 26609. National Bureau of Economic Research.

Hollander, Stephan, Maarten Pronk, and Erik Roelofsen (2010). "Does silence speak? An empirical analysis of disclosure choices during conference calls." In: *Journal of Accounting Research* 48.3, pp. 531–563.

Hong, Harrison, G Andrew Karolyi, and José A Scheinkman (2020). "Climate finance." In: *The Review of Financial Studies* 33.3, pp. 1011–1023.

Horváth, Roman et al. (2024). "Do hurricanes cause storm on the stock market? The case of US energy companies." In: *International Review of Financial Analysis*, p. 103816. ISSN: 1057-5219.

Hub, BIS Innovation (2024). *Project Gaia Enabling climate risk analysis using generative AI. BIS*.

Huynh, Thanh D and Ying Xia (2023). "Panic selling when disaster strikes: Evidence in the bond and stock markets." In: *Management Science* 69.12, pp. 7448–7467.

Jamilov, Rustam, Hélène Rey, and Ahmed Tahoun (2021). *The anatomy of cyber risk*. Working Paper 28906. National Bureau of Economic Research.

Jeon, Yuntae, Minsoo Park, Seunghee Park, et al. (2024). "GPT-based Logic Reasoning for Hazard Identification in Construction Site using CCTV Data." In: *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*. Vol. 41. IAARC Publications, pp. 291–298.

Kikstra, Jarmo S et al. (2022). "The IPCC Sixth Assessment Report WGIII climate assessment of mitigation pathways: from emissions to global temperatures." In: *Geoscientific Model Development* 15.24, pp. 9075–9109.

Koelbel, Julian et al. (2020). "Ask BERT: How Regulatory Disclosure of Transition and Physical Climate Risks affects the CDS Term Structure." In: *Swiss Finance Institute Research Paper, Journal of Financial Econometrics, forthcoming* 21.19.

Kruttli, Mathias S, Brigitte Roth Tran, and Sumudu W Watugala (2023). "Pricing Poseidon: Extreme weather uncertainty and firm return dynamics." In: *Available at SSRN 3284517*.

Lancaster, FW and EG Fayen (1973). "Information Retrieval On-Line, Melville Publ." In: *Co., Los Angeles, Calif*, p. 15.

Landsea, Christopher W and James L Franklin (2013). "Atlantic hurricane database uncertainty and presentation of a new database format." In: *Monthly Weather Review* 141.10, pp. 3576–3592.

Leippold, Markus (2023). "Sentiment spin: Attacking financial sentiment with GPT-3." In: *Finance Research Letters* 55, p. 103957.

Lenzen, Manfred et al. (2019). "Economic damage and spillovers from a tropical cyclone." In: *Natural Hazards and Earth System Sciences* 19.1, pp. 137–151.

Li, Jiazheng et al. (2020). "MAEC: A Multimodal Aligned Earnings Conference Call Dataset for Financial Risk Prediction." In: *Proceedings of the 29th ACM International Conference on Information &; Knowledge Management*. CIKM '20. Virtual Event, Ireland: Association for Computing Machinery, pp. 3063–3070. ISBN: 9781450368599. DOI: 10.1145/3340531.3412879. URL: https://doi.org/10.1145/3340531.3412879.

Li, Qing et al. (2020). "Corporate Climate Risk: Measurements and Responses." In: *Review of Financial Studies, Forthcoming*.

Loughran, Tim and Bill McDonald (2011). "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." In: *The Journal of finance* 66.1, pp. 35–65.

Mandel, Antoine, Stefano Battiston, and Irene Monasterolo (2025). "Mapping global financial risks under climate change." In: *Nature Climate Change*, pp. 1–6.

Manning, Christopher D (2009). *An introduction to information retrieval*. Cambridge university press.

Marvin, Ggaliwango et al. (2023). "Prompt engineering in large language models." In: *International conference on data intelligence and cognitive informatics*. Springer, pp. 387–402.

Mikolov, Tomas et al. (2013). "Distributed representations of words and phrases and their compositionality." In: *Advances in neural information processing systems* 26.

Moreno, Ángel Iván and Teresa Caminero (2022). "Analysis of ESG disclosures in pillar 3 reports. A text mining approach." In: *A Text Mining Approach (March 8, 2022). Banco de Espana Occasional Paper* 2204.

Moreno, Angel-Ivan and Teresa Caminero (2022). "Application of text mining to the analysis of climate-related disclosures." In: *International Review of Financial Analysis* 83, p. 102307.

Morris, Jane and Graeme Hirst (1991). "Lexical cohesion computed by thesaural relations as an indicator of the structure of text." In: *Computational linguistics* 17.1, pp. 21–48.

NGFS (2024). "Acute physical impacts from climate change and monetary policy." In: *Network for Greening the Financial System REPORT*.

Oliveira, Mariana S, Adriano Mourthe, and Maria Clara Duque (2022). "Extracting events from daily drilling reports using fuzzy string matching." In: *The APPEA Journal* 62.2, S158–S161.

OpenAI (2022). *Introducing ChatGPT*. https://openai.com/index/chatgpt/.

Qiu, Qinjun et al. (2023). "Integrating NLP and Ontology Matching into a Unified System for Automated Information Extraction from Geological Hazard Reports." In: *Journal of Earth Science* 34.5, pp. 1433–1446.

Ranger, Nicola Ann, Olivier Mahul, and Irene Monasterolo (2022). "Assessing financial risks from physical climate shocks." In: *Washington, DC: World Bank*.

Raschka, Sebastian (2014). "An overview of general performance metrics of binary classifier systems." In: *arXiv preprint arXiv:1410.5330*.

Ruas, Terry, Charles Henrique Porto Ferreira, et al. (2020). "Enhanced word embeddings using multi-semantic representation through lexical chains." In: *Information Sciences* 532, pp. 16–32.

Ruas, Terry and William Grosky (2017). "Keyword extraction through contextual semantic analysis of documents." In: *Proceedings of the 9th International Conference on Management of Digital EcoSystems*, pp. 150–156.

Sautner, Zacharias et al. (2023). "Firm-level climate change exposure." In: *The Journal of Finance* 78.3, pp. 1449–1498.

"TIGER" (2023). In: Date accessed: [07 10, 2024]: Line Shapefiles (machine-readable data files) - U.S. Census Bureau.

Van Buuren, Stef and Catharina GM Oudshoorn (2000). *Multivariate imputation by chained equations.*

Webersinke, Nicolas et al. (2021). "Climatebert: A pretrained language model for climate-related text." In: *arXiv preprint arXiv:2110.12010*.

Wei, Jason et al. (2022). "Emergent abilities of large language models." In: *arXiv preprint arXiv:2206.07682*.

Wu, Di (2024). "Text-based measure of supply chain risk exposure." In: *Management Science* 70.7, pp. 4781–4801.