2024 IEEE Conference on Artificial Intelligence (CAI)

# Optimizing Supply Chain Risk Management: An Integrated Framework Leveraging Large Language Models

Ming Zhao
*School of Business*
*UNSW Canberra*
Canberra, Australia
ming.zhao@unsw.edu.au

Omar Hussain
*School of Business*
*UNSW Canberra*
Canberra, Australia
o.hussain@unsw.edu.au

Yu Zhang
*School of Business*
*UNSW Canberra*
Canberra, Australia
m.yuzhang@unsw.edu.au

Morteza Saberi
*School of Computer Science*
*University of Technology Sydney*
Sydney, Australia
morteza.saberi@uts.edu.au

***Abstract*—The integration of Large Language Models (LLMs) in Supply Chain Risk Management (SCRM) is a novel approach to addressing the dynamic and complex challenges of risk identification and categorization in supply chains. This paper introduces a framework that leverages the capabilities of LLMs in automating the risk identification process from news and supplier databases. It also integrates a risk labeling process using the Cambridge Taxonomy of Business Risks (CTBR). A case study involving Apple Inc. as the focal company illustrates the practical application of this framework. Our methodology demonstrates significant efficiency in identifying and categorizing supply chain risks, offering a promising tool for supply chain professionals to enhance resilience and responsiveness in a rapidly evolving risk landscape.**

***Index Terms*—Supply Chain Risk Management, Large Language Models, Artificial Intelligence in Supply Chain, Risk Identification, Automated Risk Assessment, Predictive Analytics in Supply Chain**

## I. Introduction

SCRM (Supply Chain Risk Management) is crucial for maintaining the resilience and stability of supply chain companies. Traditionally, SCRM involves significant investments in human resources and time to assess information sources like news and identify potential risks. Given the high frequency of information updates and the extensive network of suppliers that supply chain companies rely on, traditional risk assessment processes often prove impractical [1], [2], [5]. The inefficiencies are primarily due to the extensive time required and the limited resources available, particularly impacting small and medium-sized enterprises (SMEs). The textual nature of risk events further complicates the feasibility of an effective SCRM.

Recent advancements in Large Language Models (LLMs) have shown promising results in both industrial and academic sectors [3], [4], [12]. In the domain of risk management, LLMs play a pivotal role in facilitating scenario-based risk evaluations by constructing various models of potential disruptions. These models cover a range of contingencies such as supplier bankruptcy, labor strikes, natural disasters, pandemics, and other similar events [14]. This paper proposes a novel LLMs-based framework for supply chain risk identification, aimed at streamlining the risk assessment process for supply chain companies. Our framework incorporates an automated process based on LLMs for identifying supply chain risks from news, complemented by a risk labeling process. This process transforms risk events described in free text into structured risk labels. The structure of the paper is as follows: Section 2 introduces our proposed framework, detailing each step; Section 3 showcases a case study involving Apple Inc., applying our framework to identify potential supply chain risks from its suppliers; Section 4 discusses the results derived from the case study in Section 3.

## II. Proposed Framework

Our framework comprises two components: an LLMs-based risk identification process and an LLMs-based risk labeling process, as shown in Figure 1. In the first risk identification process, textual inputs (e.g. news) are analysed by LLMs following our pre-defined prompts to identify potential risk events. These risk events are then be categorised to a CTBR risk label in the risk labeling process.

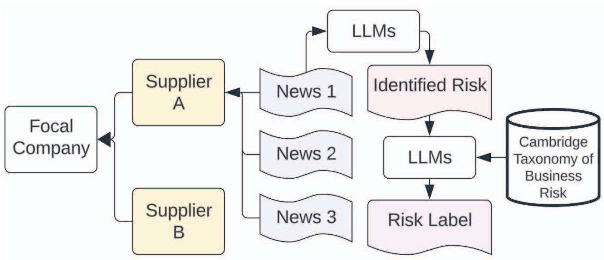


Fig. 1. Diagram of the Proposed Framework for LLMs-based SCRM

### A. Risk Identification Process

Our proposed framework depends on the availability of data from the supply chain network, complemented by continuous news data collection. It requires two key data sources: a news database and a supplier database. This data is fed into the

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00192

LLMs, which then predicts potential risk events based on news information.

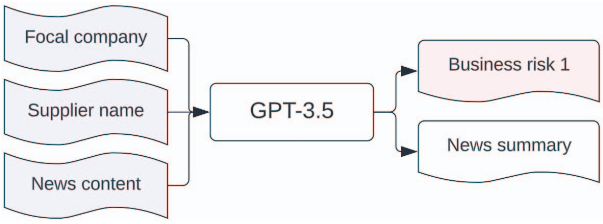


Fig. 2. Diagram of the Proposed Framework for the Risk Identification Process

In this process, we employ the GPT-3.5 (*gpt-3.5-turbo-1106*) model, a state-of-the-art language model developed by OpenAI. GPT-3.5 is characterized by its outstanding proficiency in both comprehending and generating natural language, exhibiting unparalleled aptitude in creating text that is both coherent and contextually appropriate across diverse domains. Utilizing a transformer-based architecture, this model demonstrates exceptional skill in deciphering intricate linguistic patterns and exhibits significant contextual acumen. The introduction of GPT-3.5 has significantly broadened the horizons in the realm of natural language processing, facilitating progress in areas such as text generation, language translation, and conversational artificial intelligence [12]. Its contribution to the evolution of deep learning and natural language processing has been critical, establishing GPT-3.5 as a pioneering model in these fields.

Prompt engineering plays a pivotal role in guiding LLMs to accurately interpret and respond to the provided input and its intended purpose [13]. Essentially, a prompt is a segment of natural language text that defines the task expected to be carried out by the AI. This can range from straightforward inquiries like “What is Fermat’s Little Theorem?” to directives such as “Write a poem about falling leaves”, and even brief feedback phrases including “too verbose”, “too formal”, “rephrase again”, “omit this word”. It may also involve detailed statements that provide context, instructions, and input data. The practice of prompt engineering is vital in refining the model’s response to align with specific goals and criteria.

The prompts we designed for the risk identification process are as follows:

1) role: system - You are a profressional risk assessor working at the focal company: {focal_company}.
2) role: system - Use the following step-by-step instructions to respond to user inputs. If the answer cannot be found in the articles, output ‘N/A’.
3) role: system - Step 1 - The user will provide you with text in triple quotes. Summarize the news content.
4) role: system - Step 2 - Determine whether the provided news is related to the supplier: {supplier_name} with the rationale.
5) role: system - Step 3 - Determine whether the provided news is related to the focal company: {focal_company} with the rationale.
6) role: system - Step 4 - If the result in the Step 2 is ‘Related’, then proceed; otherwise output ‘N/A’ for all following steps.
7) role: system - Step 5 - Identify potential business risks related to the supplier: {supplier_name}.
8) role: user - Supplier: {supplier_name} News content: {news_content}

To obtain structured results from the GPT-3.5 generative AI model, we have incorporated function calling within our API interactions. This approach allows users to define functions in their API requests, enabling the model to intelligently generate a JSON object containing arguments for invoking one or more functions. We designed a specific JSON object structure to outline the expected outputs, detailing each property, including variable names, data types, and descriptions. In conjunction with the custom-designed prompts, this setup ensures that each API call to the GPT-3.5 model returns a JSON string. This string can then be easily parsed to extract results for each defined field, facilitating a structured and efficient data retrieval process.

In our methodology, we calibrated the generative process by setting the temperature parameter to 0. This adjustment was essential to control the randomness in the text output, ensuring a consistent and predictable generation of content for each query. Furthermore, we established a static seed, a crucial step to guarantee uniformity in the results across different queries. This approach ensures that each query produces a consistent generative output, a key aspect for comparative analysis and reproducibility in our research.

### *B. Risk Labeling Process*

After identifying risk events from news sources using the GPT-3.5 model, each event is recorded in a textual format. However, managing and querying these text-based risk events becomes challenging, given the sheer number of suppliers and the frequency of news updates. To address this, we have integrated a system with the Cambridge Taxonomy of Business Risks (CTBR) to classify these events into structured business risk labels.

CTBR is a comprehensive taxonomy that encompasses a wide range of potential threats to businesses [15]. In the CTBR, risks are categorized based on their cause and are organized into a hierarchical structure defined by Class, Family, and Type. This structure is based on clustering principles that emphasize similarity and commonality. While many risk types could logically fit into several families, and there could be debates over the allocation of individual risks to different classes, CTBR has assigned each risk type to a single category to maintain clarity and avoid redundancy. This structured approach enhances the feasibility of managing and querying risk events, making it a more practical solution for businesses dealing with a vast array of risks.

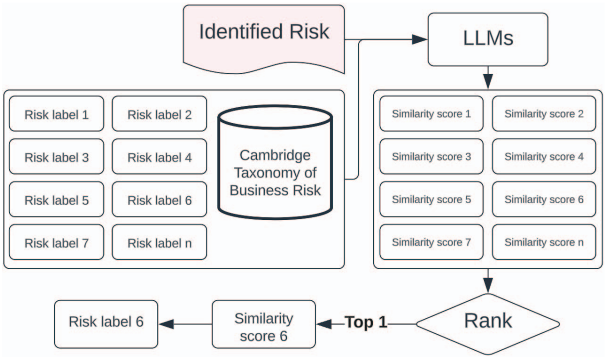


Fig. 3. Diagram of the Proposed Framework for the Risk Labeling Process

In the CTBR, business risks are categorized into six classes: Financial, Geopolitical, Technology, Environment, Social, and Governance. Within each risk class, there are several risk families. For example, under the Financial risk class, families such as Economic Outlook and Economic Variables are included. Further, under these risk families, there are specific risk types. For instance, Recession and Stagnation fall under the Economic Outlook risk family. Therefore, a typical risk item in the CTBR follows the format: Risk Class - Risk Family - Risk Type, such as Financial - Economic Outlook - Recession.

In addition to the labels of risk classes, families, and types, CTBR also provides definitions for each category. These definitions give further semantic clarity. For instance, the Financial risk class is defined as: "Threats from the macroeconomy, financial markets, global economic value chains, industry or company-specific events leading to underperformance of corporates."

To categorize each risk event into a predefined risk label (Risk Class - Risk Family - Risk Type), we employ semantic text similarity analysis between the risk event and the risk label. This process involves the use of embeddings, which are vector representations capturing the semantic information of words, phrases, or entire texts in a continuous vector space. These embeddings are usually derived from neural network-based models like Word2Vec, GloVe, or BERT. They encode contextual and semantic relationships between words, enabling the calculation of text similarity. This similarity is often measured using metrics such as cosine similarity or Euclidean distance. By converting text into high-dimensional vector representations, embeddings capture the nuances of language and context, allowing for precise assessment of semantic relatedness and similarity. This methodology is widely used in various natural language processing tasks, including information retrieval, document clustering, and recommendation systems.

Considering the limitations of our computing resources, we evaluated six base sized popular open-source pre-trained generative language models, all of which are available for download from HuggingFace. These models were selected for their efficacy in generating embeddings and their suitability for our text similarity assessment needs.

- bert-base-uncased[1] [6]
- all-mpnet-base-v2[2] [7]
- sentence-t5-base[3] [8]
- all-MiniLM-L6-v2[4] [9]
- gtr-t5-base[5] [10]
- all-distilroberta-v1[6] [11]

For categorizing risk events into specific risk labels from the CTBR, we employ a process where both the risk event and the risk labels are fed into pre-selected models to calculate their semantic similarity. The most appropriate risk label for a given event is determined based on the highest similarity score obtained from this comparison.

We designed the following 4 approaches combining different parts from the CTBR to be used to calculate semantic text similarity:

1) Risk Class + Risk Family + Risk Type
2) Risk Class + Risk Class Definition + Risk Family + Risk Type
3) Risk Class + Risk Class Definition + Risk Family + Risk Family Definition + Risk Type
4) Risk Class + Risk Class Definition + Risk Family + Risk Family Definition + Risk Type + Risk Type Definition

Cosine similarity approach is involved to convert the text of both the risk event and the risk label into vector representations using embeddings. Once in vector form, we calculate the cosine similarity between these vectors. The cosine similarity metric measures the cosine of the angle between two vectors, effectively gauging their orientation in the vector space. A higher cosine similarity indicates a greater degree of similarity between the two text segments.

After calculating the cosine similarities for all risk labels within the CTBR, a ranking process is employed to identify the risk label most closely aligned with the specified risk event. This most proximate risk label is then subsequently assigned to the risk event in question.

## III. A Case Study

In this case study, we illustrate the application of our proposed framework using Apple Inc. as the focal company. In Supply Chain Risk Management (SCRM), the "focal company" is typically the central entity in a supply chain network, often playing a crucial role in coordinating and managing activities across the supply chain.

### A. Dataset

For this study, we sourced the list of Apple's tier-1 suppliers from publicly available information[7]. This list accounts for

[1] https://huggingface.co/bert-base-uncased
[2] https://huggingface.co/sentence-transformers/all-mpnet-base-v2
[3] https://huggingface.co/sentence-transformers/sentence-t5-base
[4] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2
[5] https://huggingface.co/sentence-transformers/gtr-t5-base
[6] https://huggingface.co/sentence-transformers/all-distilroberta-v1
[7] https://www.apple.com/supplier-responsibility/pdf/Apple-Supplier-List.pdf

98% of Apple's direct expenditures on materials, manufacturing, and assembly of its products globally for the fiscal year 2022. It includes 188 distinct suppliers, along with their names and primary locations.

In addition to the supplier list, we compiled a dataset of news articles related to these suppliers. This dataset was curated using Google News and includes a range of associated news items. Each news article in the dataset contains the raw HTML content and relevant metadata such as the article's title, the date and time of publication, and the name of the media outlet. For the purpose of this case study, we restricted our dataset to news articles published in the year 2023.

This comprehensive data collection provides a rich foundation for applying our risk identification framework. By analyzing the news related to Apple's suppliers, we can identify and categorize risk events using our previously outlined methodologies. This approach enables us to demonstrate the practical application and effectiveness of our framework in a real-world context, specifically in managing and mitigating risks in a complex and dynamic supply chain like Apple's.

### B. Risk Identification Process

In our case study, we processed news articles from Apple's supplier news database using the GPT-3.5 model to generate associated risk events. Each news article was analyzed following the previously mentioned prompts. Previous experiments have proved the capacity of directly extracting HTML content in the GPT-3.5 model. A sample identified risk event is as follows: "The sale of the equity stake may impact 3M's financial results and business prospects. It may also affect the availability and cost of purchased components, compounds, raw materials, and energy due to supply chain interruptions."[8]

We constructed a evaluation dataset consisting of 115 risk events identified from the risk identificaiton process from 21 news articles.

### C. Risk Labeling Process

Once these risk events were identified from news about Apple's suppliers, they were categorized using the risk labels defined in the CTBR. The CTBR comprises a comprehensive set of 875 individual risk labels, encompassing all risk classes, families, and types. Each identified risk event from the news was then evaluated for semantic similarity with each of the CTBR risk labels.

Given that all risk events in this study were generated by the LLM and not derived from an existing benchmark dataset, there were no ground truth values to measure the accuracy of the risk labeling process. Consequently, we undertook a manual labeling and evaluation process for 115 of these risk events, comparing them with their predicted outputs from the six LLMs. This labeling process involved three experts, each of whom independently labeled the 115 risk events against all associated risk labels from the CTBR.

[8]https://www.prnewswire.com/news-releases/3m-announces-senior-management-team-changes-301806620.html

The results of the correctly labeled risk events are presented in Table I. For assessing the accuracy of our risk labeling, we employed three different types of counting methods for the correctly labeled risk events, which are as follows:

- Any: A label is deemed accurate and marked as true if at least one expert identifies it as the correct label.
- 2/3: A label is classified as true if the majority of experts, specifically two-thirds, concur on its correctness.
- All: A label is confirmed as true only if all experts unanimously agree on its accuracy.

TABLE I
RISK LABELING PERFORMANCE BENCHMARK

| model | type | cfr[a] | cfr_rd[b] | cfr_fdrd[c] | cfr_cdfdrd[d] |
|---|---|---|---|---|---|
| all-distilroberta-v1 | any | 57 | 52 | 62 | 66 |
| gtr-t5-base | any | 66 | 66 | 61 | 63 |
| bert-base-uncased | any | 38 | 42 | 54 | 52 |
| all-mpnet-base-v2 | any | 67 | 63 | 57 | 62 |
| all-MiniLM-L6-v2 | any | 57 | 63 | 59 | 69 |
| sentence-t5-base | any | 77 | 77 | 79 | 77 |
| all-distilroberta-v1 | 2/3 | 47 | 42 | 48 | 51 |
| gtr-t5-base | 2/3 | 59 | 53 | 52 | 54 |
| bert-base-uncased | 2/3 | 30 | 26 | 39 | 41 |
| all-mpnet-base-v2 | 2/3 | 56 | 49 | 45 | 44 |
| all-MiniLM-L6-v2 | 2/3 | 50 | 52 | 52 | 59 |
| sentence-t5-base | 2/3 | 68 | 64 | 63 | 62 |
| all-distilroberta-v1 | all | 35 | 31 | 38 | 37 |
| gtr-t5-base | all | 51 | 41 | 42 | 45 |
| bert-base-uncased | all | 18 | 17 | 26 | 30 |
| all-mpnet-base-v2 | all | 39 | 33 | 34 | 32 |
| all-MiniLM-L6-v2 | all | 42 | 39 | 40 | 41 |
| sentence-t5-base | all | 54 | 46 | 51 | 51 |

**a** cfr: includes Risk Class + Risk Family + Risk Type.
**b** cfr_rd: includes Risk Class + Risk Class Definition + Risk Family + Risk Type.
**c** cfr_fdrd: includes Risk Class + Risk Class Definition + Risk Family + Risk Family Definition + Risk Type.
**d** cfr_cdfdrd: includes Risk Class + Risk Class Definition + Risk Family + Risk Family Definition + Risk Type + Risk Type Definition.

Based on the results shown in Table I, the sentence-t5-base model achieved the highest overall accuracy and has been selected as the candidate model for the risk labeling process in our framework.

An interesting observation from our analysis is that more complex LLMs, such as sentence-t5-base, do not necessarily require the explicit inclusion of additional definitions to yield accurate results. This model demonstrated a relatively higher degree of accuracy even without such enhancements. In contrast, for less complex models like BERT, the inclusion of more detailed definitions significantly improved performance.

In Figure 4, we illustrate an example workflow for Apple, focusing on its supplier 3M. We selected a specific news item related to 3M and processed it through our framework. During the risk identification stage, one associated risk event was identified from this news article. Subsequently, in the risk labeling process, this identified risk event was assigned a corresponding risk label from the CTBR. This allocation was determined based on the ranking derived from the semantic text similarity analysis of each label.

This workflow exemplifies how our proposed framework can be applied in a real-world scenario, demonstrating its effectiveness in identifying and categorizing risk events within a complex supply chain environment. By automating the process of risk event identification and labeling, our framework enhances the efficiency and accuracy of supply chain risk management, particularly for large organizations with extensive supply networks like Apple.

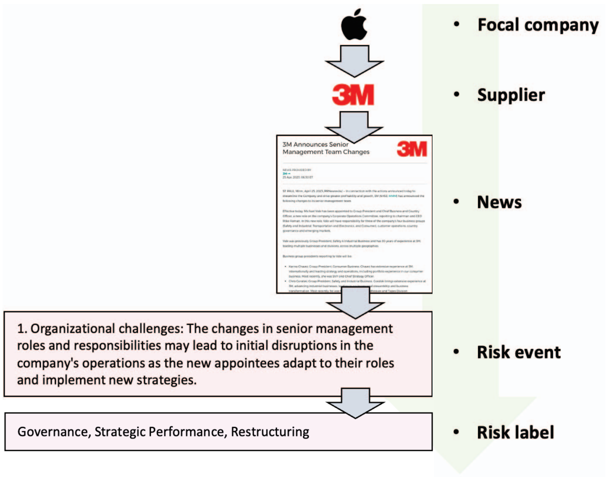


Fig. 4. Workflow Example Featuring Apple Inc. and a News of 3M

## IV. Results

Our framework's application in the case study with Apple Inc. yielded promising results. From the evaluation dataset of 115 risk events identified across 21 news articles, our LLM-based approach demonstrated high efficiency in detecting and categorizing supply chain risks. The integration of CTBR further improved the organization and clarity of risk categorization, leading to more actionable insights for risk management. The *sentence-t5-base* model showcased the highest accuracy in the risk labeling process, outperforming other models in various configurations. This framework, incorporating LLMs, underscores the transformative potential of advanced LLMs in enhancing SCRM.

As a risk manager within the focal company, our proposed framework provides a streamlined method for selecting specific types of risks through CTBR. This framework integrates the processes of risk identification and labeling, efficiently linking identified risk events with their corresponding risk labels, and further includes associated news sources and suppliers for reference. Such integration ensures that risk managers receive a comprehensive overview of relevant risks, tailored to their needs. Figure 5 illustrates how our framework facilitates an intuitive and informative presentation of risks to risk managers. This enhancement aids them in making informed decisions and managing risks effectively.

Figure 6 presents an interactive result of our proposed framework. Upon selection of risk labels from the CTBR by

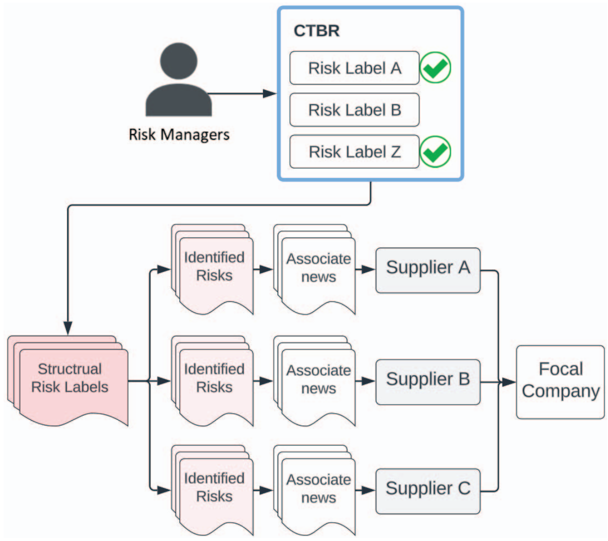


Fig. 5. End-User Interaction with the Framework

risk managers, the system dynamically presents all associated risk events that have been identified through the LLMs. For each risk event, detailed information including reference to the corresponding supplier and the original news sources are also displayed. This feature not only enhances the usability of our framework but also provides a comprehensive and insightful view of the risk landscape, tailored to the specific needs and preferences of the user.

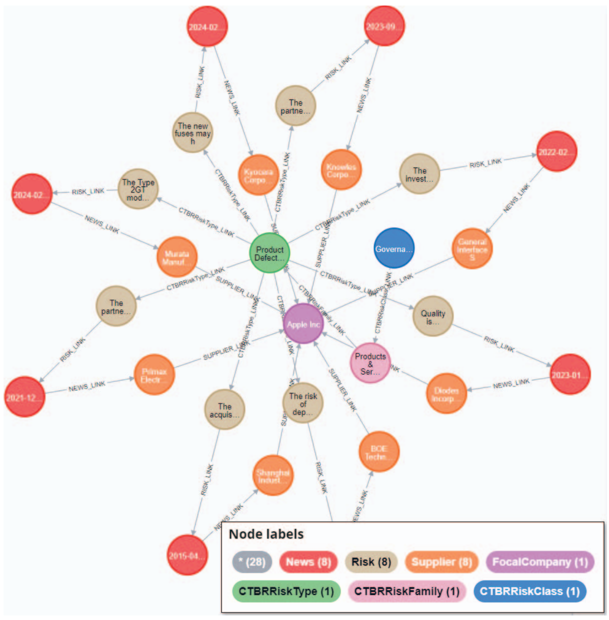


Fig. 6. Graph Displaying Result Filtered by a Specific Risk Label

The comparative analysis of different LLMs revealed interesting findings. Models like sentence-t5-base were able to achieve high accuracy even without additional contextual information, while others benefitted significantly from the inclusion of detailed definitions. This underscores the importance of model selection based on the specific requirements of the SCRM process.

In summary, the proposed LLM-based framework demonstrates a high potential for improving the efficiency and accuracy of risk identification and categorization in supply chains. Its application in a real-world scenario with Apple Inc. validates its practicality and effectiveness, marking a significant advancement in the field of SCRM.

## V. Summary and Future Works

In this paper, we present our ongoing efforts in integrating LLMs into SCRM. LLMs have demonstrated their effectiveness across various domains. In this study, we address the challenges posed by SCRM's dependency on frequently updated and extensive information sources, such as news, coupled with the constraints of limited human resources and time. Traditional methods in SCRM tend to focus on specific domain knowledge or rule-based approaches, which restrict their applicability across different sectors or industries. However, the recent advancements in LLMs offer a solution: General pre-trained LLMs can now be adapted for use in diverse sectors, enabling the prediction of risk events with reasonable accuracy. Furthermore, by integrating a structured taxonomy, the free-text risk event data generated by LLMs can be effectively organized and made accessible for end-user interpretation and application.

The future work for this study involves several key areas of expansion and development:

- A continuous effort will be made to evaluate more Large Language Models, especially those of larger scale such as GPT-4. Additionally, open LLMs like Bloom-176b[9] and Llama-2-70b[10] will be assessed if possible. The inclusion of these advanced models could offer deeper insights and potentially more accurate risk identification and categorization due to their expansive training datasets and advanced architectures.
- A software prototype tailored for supply chain end-users is currently in development. This software is designed to facilitate the practical application of our framework in real-world supply chain management scenarios. The prototype aims to be user-friendly and adaptable, catering to the specific needs and dynamics of different supply chains.
- Future studies will include more case studies involving a variety of focal companies and their suppliers. This will allow us to test the efficacy and adaptability of our framework across different industries and supply chain structures. By doing so, we aim to validate the universality of our approach and refine it to cater to the diverse challenges and risk profiles encountered in different supply chain contexts.

These future endeavors are aimed at enhancing the robustness, accuracy, and applicability of our framework in the evolving landscape of SCRM. By incorporating more advanced models and broadening the scope of our case studies, we hope to offer a more comprehensive tool for risk assessment and mitigation in supply chains, ultimately contributing to more resilient and efficient supply network management.

[9]https://huggingface.co/bigscience/bloom
[10]https://huggingface.co/meta-llama

## References


[1] W. Ho, T. Zheng, H. Yildiz, and S. Talluri, "Supply chain risk management: a literature review," International journal of production research, vol. 53, no. 16, pp. 5031-5069, 2015.

[2] M. S. Sodhi and C. S. Tang, Managing supply chain risk. Springer Science & Business Media, 2012.

[3] T. Brown et al., "Language models are few-shot learners," Advances in neural information processing systems, vol. 33, pp. 1877-1901, 2020.

[4] L. Floridi and M. Chiriatti, "GPT-3: Its nature, scope, limits, and consequences," Minds and Machines, vol. 30, pp. 681-694, 2020.

[5] G. Baryannis, S. Validi, S. Dani, and G. Antoniou, "Supply chain risk management and artificial intelligence: state of the art and future research directions," International Journal of Production Research, vol. 57, no. 7, pp. 2179-2202, 2019.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[7] K. Song, X. Tan, T. Qin, J. Lu, and T.-Y. Liu, "Mpnet: Masked and permuted pre-training for language understanding," Advances in Neural Information Processing Systems, vol. 33, pp. 16857-16867, 2020.

[8] J. Ni et al., "Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models," arXiv preprint arXiv:2108.08877, 2021.

[9] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers," Advances in Neural Information Processing Systems, vol. 33, pp. 5776-5788, 2020.

[10] J. Ni et al., "Large dual encoders are generalizable retrievers," arXiv preprint arXiv:2112.07899, 2021.

[11] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," arXiv preprint arXiv:1910.01108, 2019.

[12] M. U. Hadi et al., Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects. 2023.

[13] J. Ye et al., "A Comprehensive Capability Analysis of GPT-3 and GPT-3.5 Series Models," ArXiv, vol. abs/2303.10420, 2023.

[14] R. G. Richey Jr., S. Chowdhury, B. Davis-Sramek, M. Giannakis, and Y. K. Dwivedi, "Artificial intelligence in logistics and supply chain management: A primer and roadmap for research," Journal of Business Logistics, vol. 44, no. 4, pp. 532-549, 2023, doi: https://doi.org/10.1111/jbl.12364.

[15] C. R. Framework, "Cambridge Taxonomy of Business Risks," 2019.