HONG KONG MONETARY AUTHORITY

Research Memorandum 07/2025

26 March 2025

UTILISING GENERATIVE AI TOOLS TO MONITOR SYSTEMIC RISKS IN GLOBAL BANKING: AN ANALYSIS OF EARNINGS CALL TRANSCRIPT DATA

Key points:

- *Generative AI (GenAI) is an effective tool for analysing large volumes of unstructured textual data. In financial stability analysis, assessing such information can complement traditional monitoring approaches that rely primarily on quantitative data, which often face issues related to data availability and time lags.*

- *To enhance our monitoring tools, this paper introduces a framework that uses our in-house large language model (LLM) to analyse banks' earnings call transcripts to uncover insights into key risks and new emerging vulnerabilities concerning the global banking industry from market participants.*

- *Our analyses of over 11,600 transcripts from 520 publicly listed banks worldwide between 2019 and 2024 indicate that the framework is capable of identifying key risk factors from these transcripts that align with actual risk drivers observed during recent industry-wide stress episodes, such as the Russia-Ukraine conflict and the US banking turmoil. By using various data visualisation tools, including network analysis, the framework enables us to examine the interconnections between key risk factors, which facilitates the monitoring of potential risk spillovers.*

- *Moreover, the framework is found to be able to provide reliable early warning signals ahead of actual deterioration in financial indicators prior to stress events. For instance, taking the stresses stemming from US banking exposures to commercial real estate (CRE) as an example, the framework is able to signal the CRE-related risks from the transcripts of US banks ahead of the actual deterioration in the delinquency ratio of US banks' CRE loans. Together, these analyses demonstrate that the framework is useful for monitoring systemic risks in global banking.*

- *Our latest assessment, based on transcripts from 2024H2, reveals that credit and geopolitical risks are two important risk factors that have recently gained increasing attention in the global banking industry. Increased discussions of these two risks* *probably reflect continued credit quality concerns under the possible "high-for-longer" interest rate environment and renewed policy uncertainties surrounding trade tariffs and foreign relations under the new US administration. Thus, closer monitoring of these two areas is warranted.*

- *Finally, although the framework is demonstrated to be useful for monitoring systemic risks, it is important to recognise the limitations of GenAI, including its 'black box' nature. Therefore, it should be applied alongside traditional monitoring tools to mitigate potential modeling risks.*

Prepared by: Andrew WONG, Kelvin HO and Icarus CHAN[1]
Market Research Division, Research Department
Hong Kong Monetary Authority

[1] The authors would like to thank Lillian Cheung and Eric Wong for their helpful comments and suggestions. We would also like to express our gratitude to Maggie Poon, Oscar So, and Max Kwong for their valuable technical support on the in-house Generative AI tools. The views expressed in this paper are those of the authors, and not necessary those of the Hong Kong Monetary Authority.

I. INTRODUCTION

Generative AI (GenAI) is a powerful and efficient tool for analysing huge amounts of unstructured textual data. GenAI is useful for monitoring risks in the global banking industry, as textual information from market participants (e.g. transcripts of earnings calls) may provide timely and important insights into key and emerging vulnerabilities. Assessing such information can complement traditional monitoring approaches that rely heavily on quantitative data, which may face issues related to data availability and time lags.

Researchers from both academia and the central bank community are increasingly adopting this new tool to gain fresh insights from textual information released by corporations. Recent studies (e.g., Andersson, Guillotin and Neves, 2024; Hollander et al., 2024; Jha et al., 2024) have demonstrated that corporate textual disclosures coupled with suitable textual processing can reveal valuable information about economic outlooks or specific industry trends (e.g. Soto, 2019; Cook et al., 2023; Masson, 2023). These timely and valuable insights can then be used for various purposes, including economic surveillance and signal generation.

Against this background, this paper introduces a framework that uses the large language modeling (LLM)[2] technique to analyse banks' earnings call transcripts, aiming to generate insights into the key and emerging risks facing the global banking industry. The framework is shown to be effective in providing important analytical insights both at the industry and bank levels. We apply the framework to a large sample of banks' earnings call transcripts between 2019 and 2024, and our findings suggest that the framework can identify key risk factors that align with actual risk drivers observed during several recent industry-wide stress episodes. In addition, the framework is found to be able to provide reliable early warning signals ahead of the actual deterioration of financial indicators prior to the stress events. This demonstrates that the framework is effective in monitoring systemic risks in global banking. Lastly, our latest assessment based on 2024 H2 data reveals that geopolitical and credit risks, two important risk factors that have recently gained increasing attention in the global banking industry, warrant closer monitoring.

This paper is structured as follows. Section II describes the sources and sample coverage of the earnings call transcript data used in this paper. Section III outlines the details of the analytical framework. Section IV provides several case studies to demonstrate how risk insights can be obtained from the framework with reference to four recent banking industry stress episodes. In Section V, we examine

[2] Accessible guidance on LLMs in the economic and finance domain can be found in Kwon et al. (2024).

the latest risk situations facing the global banking industry based on the assessment of available earnings call transcripts in 2024H2. The final section concludes the paper.

II. DATA DESCRIPTION

Banks' earnings call transcripts are sourced from the S&P Capital IQ and Capital Pro databases.[3] These transcripts record all of the discussion content of call events and capture various metadata, such as dates, company names, speakers, and call content for publicly listed banks, between 2019 and 2024.[4] An earnings call conference, which typically consists of a presentation section and a question and answer (Q&A) section, is an important event for the management of a company to engage with their investors and analysts. These transcripts thus contain timely and rich qualitative information about a bank's current business performance, outlooks, and risk concerns expressed by the bank's management, market analysts, and investors.

While earnings call conference events are typically held quarterly in the US, the frequency of such evens for banks outside the US is lower (e.g., semi-annually or annually). To balance the geographical scope of the banks in the sample and the degree of timeliness, we apply our framework to a sample of English transcripts on a semi-annual basis (e.g., 2019H1, 2019H2). Because earnings calls may be conducted in languages other than English (particularly for banks in the Asia-Pacific region), we also include machine-translated English transcripts to enhance the representativeness of the listed banks in our analysis. We exclude all transcripts of calls that are not held to discuss the earnings performance or operating performance of a bank (e.g., fixed income calls, special calls, merge and acquisition calls, sales and trading calls).

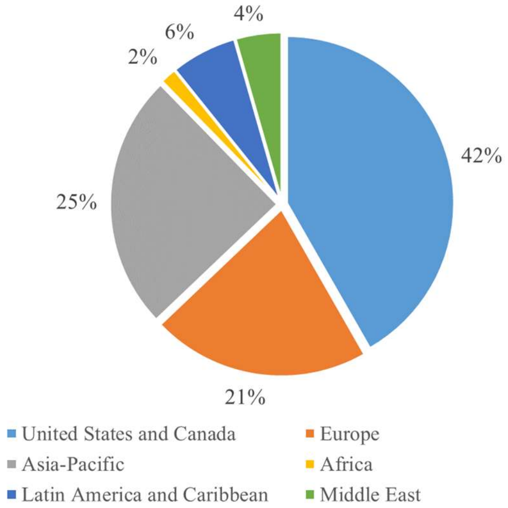
Over 11,600 transcripts from 520 publicly listed banks are used in this study, of which around 42%, 25%, and 21% are located in North America, Asia-Pacific, and Europe, respectively (Figure 1). Additional descriptive information, such as primary industry, countries and regions of headquarters, and other bank-specific metadata for individual banks, is obtained from the S&P Capital IQ Pro database.

---

[3] As the raw transcripts are available in PDF format, we follow the approach adopted by Li et al. (2021a, 2021b) and extract the relevant contents from pdf files into CSV format. The relevant Python codes for extracting the content are available for download from Github from the authors (https://github.com/ssrn3632395/The-Role-of-Corporate-Culture-in-Bad-Times). We express our gratitude to Li et al. (2021a) for their generosity in uploading the open-source codes online for others' reference.

[4] Although transcripts are also available for many banks before 2018, the coverage of transcripts from banks located in regions outside the US and Canada is more limited between 2012 and 2018 compared with the period after.

Figure 1: Distribution of banks by geographical markets

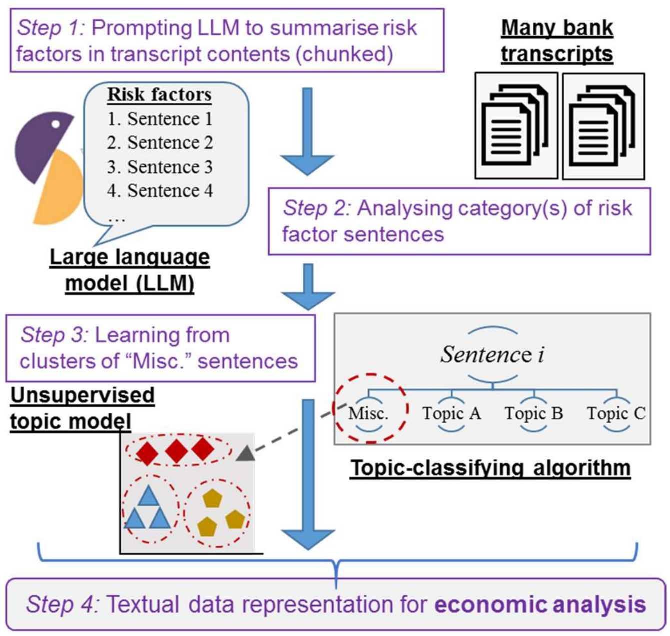


Source: HKMA staff estimates based on S&P Capital IQ transcript data and the S&P Capital IQ Pro database.

III. OVERVIEW OF THE ANALYTICAL FRAMEWORK

In this subsection, we outline the details of our analytical framework for analysing the transcript data. Figure 2 displays a diagram that summarises the key steps of the workflow process. This process can be mainly categorised into four steps: text summarisation, topic classification, topic learning, and quantitative analysis. Each step is discussed in more detail in the following.

Figure 2: Diagram of the framework



Note: "Misc." indicates the group of risk factor sentences that the topic-classifying algorithm assigns to the "miscellaneous" group.

Step 1: Summarising risk discussions from the transcript data

In the first step, the transcripts are processed by an LLM to extract the relevant risk discussions. We use our in-house GenAI tool (Secured AI Research Assistant, "SARA") to condense lengthy transcripts into shorter, more structured, and more readable formats, while retaining the key information.[5]

To guide the LLM in generating the outputs in the desirable format, a prompt template is designed (see Figure 3) to instruct the LLM to generate up to seven sentences about the downside risks of the company that have been discussed based on a provided section from the transcript.[6] These outputs (i.e., summarised risk sentences) facilitate analysis in subsequent steps. Importantly, given the structured output format, it is feasible to trace the source of each summarised risk sentence back to the corresponding presentation and Q&A section of each transcript, which enables further in-depth analysis. An example of this summarisation process is provided in Figure A.2 of Appendix VII.

---

[5] SARA is a GenAI tool that is built upon a wide range of open-source LLMs, including the general-purpose Llama3-70B model (Grattafiori et al., 2024) developed by Meta Inc. Although more advanced English-based LLMs are available in the market (e.g., the GPT series and the Claude series of models), we did not select them for this analysis because of their closed-source nature, which makes it challenging for researchers to fully comprehend their data sources. Supplementary details about the hyper-parameter setting of our use case with the Llama3-70B model are provided in Appendix I.

[6] We follow the practice in the literature of chunking the contents of transcripts into several subsections for LLM processing. Details of the text preprocessing are available in Appendix II.

Figure 3: Prompt template and output structure

Upper panel: Prompt template
System: You are a helpful financial analyst responsible for analyzing financial disclosure documents. You follow instructions extremely well. You work on an extractive summarization task and your task is to find whether there is any potential downside risk related to the company and then advise clients. Read the given text from an earnings call transcript. Your task is to determine whether there is any negative factor about the company discussed. If there are any negative factors discussed, extract and summarize at most seven negative factors in bullet points, your output should be structured like this: ASSISTANT: Negative factors: 1. ... 2. ... 3. ... 4. ... 5. ... 6. ... 7. ... Or, if there is no negative factor discussed, your output should be structured the same as following: ASSISTANT: Negative factors: 1. Nil {INSERT PASSAGE}

Bottom panel: Output structure format by the LLM

(i) Hypothetical case A: 4 risk factors were found in the passage:	(ii) Hypothetical case B: No risk factor is found in the passage:
Output: Negative factors: 1. {SENTENCE A} 2. {SENTENCE B} 3. {SENTENCE C} 4. {SENTENCE D}	Output: Negative factors: 1. Nil

Step 2: Classifying risk sentences into topics relevant to the banking industry

In the second step, the summarised risk sentences are classified into 37 risk topics based on a text classification algorithm specific to the banking industry. This enables us to measure the frequencies of the different risk topics mentioned in an earnings call transcript.

The 37 pre-defined risk topics are derived based on a comprehensive review of the literature and financial stability reports, as we aim to capture the major risk factors that are commonly cited in the global banking industry.[7] Detailed information about the 37 risk topics is provided in Table A.4 of Appendix VII. These topics can be grouped into 13 broader categories, such as earnings risk, interest rate risk, credit risk, macroeconomic risk, operational and compliance risk, and business growth risk, for easier interpretation.

A classification algorithm based on the sentence transformer model[8] (Reimers and Gurevych, 2019) is used to systemically classify the summarised risk sentences into these 37 risk topics. Figure 4 presents a diagram of the classification algorithm, which involves three main steps. First, each of the 37 risk topics is tagged with a representative sentence that serves as the anchor sentence for each risk topic. Second, a fine-tuned application[9] of the sentence transformer model is used to generate the embedding vectors for both the summarised risk sentences and the anchor sentences associated with each of the 37 risk topics. A cosine similarity score[10] is then computed between the embedding vectors of the summarised risk sentences and each of the anchor sentences, which forms the basis for risk topic classification. If the similarity score for a summarised risk sentence and an anchor sentence exceeds a threshold of 0.6, the risk sentence is classified under that specific risk topic.[11]

---

[7] These include financial stability reports and banking supervisory reports published by international organisations, such as the Bank for International Settlements, the Basel Committee on Banking Supervision, the Financial Stability Board, and the International Monetary Fund. We also refer to financial stability reports from jurisdictional supervisory bodies such as the Federal Reserve Board, the Bank of England, the European Central Bank Banking Supervision, and the Hong Kong Monetary Authority. Earlier research papers, including Soto (2019), Wei et al. (2019), Cook et al. (2023), and Masson (2023), are also reviewed to refine the list of risk topics.

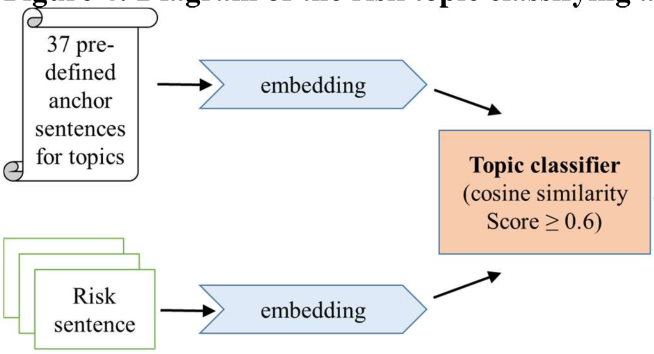
[8] The sentence transformer is a modification of the Bidirectional Encoder Representations from Transformers architecture for applications in generating semantically meaningful sentence embedding based on a pre-trained word embedding model.

[9] One shortcoming that may arise when applying pre-trained word embedding models in the finance domain is that models are typically trained on huge amounts of publicly available data, and therefore may not have sufficient understanding of banking industry terminology. To address this, we conduct additional fine-tuning on the word embedding model with more than 300 labeled risk sentences for the 37 pre-defined topics. Details are provided in Appendix III.

[10] For any risk sentence SENTi, the pair-wise cosine similarity score with the pre-defined topic j using the AnchorSENTj can be derived as $CSscore(SENT_i, AnchorSENT_j) = \frac{\|SENT_i \cdot AnchorSENT_j\|}{\|SENT_i\|\|AnchorSEN\ _j\|}$, where $\|any\ sentence\|$ is the embedding vector representation of a sentence based on the selected word embedding model under the sentence transformer architecture. A higher CSscore reflects a closer directional angle between the embedding vectors of the two sentences, which should imply a higher degree of commonality in the content covered between the two sentences.

[11] In practice, cosine similarity score values between 0.6 and 1 are often considered to be strong indicators of similarity (Manralai, 2023).

Figure 4: Diagram of the risk topic classifying algorithm



Topic	Score (sorted)	Topic assigned
G	0.78	Y
B	0.63	Y
E	0.4	N
F	0.2	N
A	0.1	N
C	0.05	N

Because each summarised risk sentence may be related to multiple topics (i.e. the risk sentence has a similarity score exceeding 0.6 for more than one topic), it can be classified into a maximum of three of the 37 pre-defined topics (sorted in descending order). For example, in Figure 4, the risk sentence is classified into topics G and B because the risk sentence has similarity scores of 0.78 and 0.63, respectively, for them. Any risk sentences that do not fit into the 37 pre-defined specific topics are classified under the "miscellaneous" category for further topic learning in the third step.

Step3: Detecting emerging topics through unsupervised learning

Bank management, analysts and investors may bring up new risks and bank-specific issues that may not be captured by the 37 pre-defined categories in the previous step, and these miscellaneous risk sentences may contain useful information about new emerging risks and vulnerabilities. To uncover insights from these miscellaneous sentences, we use the BERTopic model (Grootendorst, 2022), which is an unsupervised topic modeling tool, to analyse and cluster them.[12] The key intuition behind this topic model is to identify clusters of sentences that share similar contexts so that meaningful labels can be learned from the characteristics of sentences within each cluster.[13] These topic labels, which can be obtained by data visualisation tools such as word cloud(s) and LLM labeling techniques, could help us to identify any new emerging risk factors facing the global banking industry. Additional details about the implementation of the BERTopic architecture can be found in Appendix IV.

Step 4: Economic analysis of textual data representations

[12] BERTopic is a topic modeling technique that leverages embedding models to capture semantic relationships within text data, enabling it to identify coherent topics from a large corpus of sentences.

[13] In applications, the BERTopic algorithm first transforms sentences into high-dimensional vectors using pre-trained word embedding models, capturing contextual information and nuances in language. It then applies a dimensionality reduction technique to project these embedding vectors into a lower-dimensional space while preserving their semantic structure. A clustering algorithm is subsequently used to group similar sentences.

The classified risk topics from all of the risk sentences (i.e. outputs from steps 2 and 3) can then be used to quantitatively assess important factors and trends, as well as to explore any interconnection between these risk factors in the global banking industry over time.

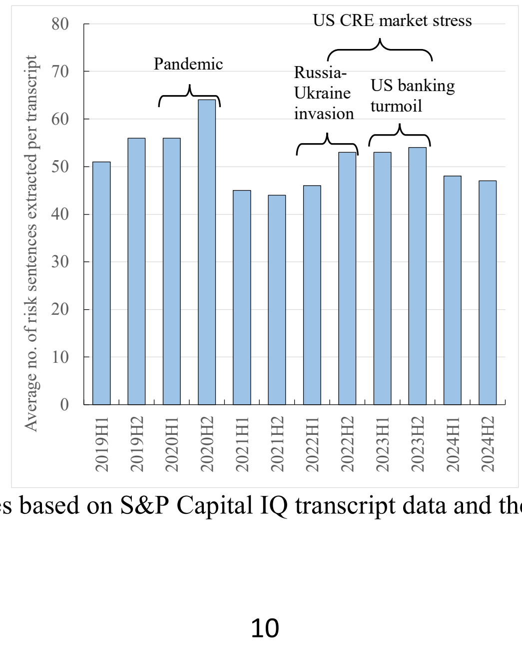
It is important to note that the number of transcripts available for analysis could vary over time for various reasons, such as if bank management decides not to hold a meeting at certain times or if the transcript records are not available in the database. To ensure comparability across time, we scale the relevant metrics by the number of transcripts in the same period.

IV. ANALYTICAL RESULTS FROM INDUSTRY-WIDE STRESS EPISODES

In this section, we provide three applications to show how risk insights can be obtained from our framework. We use various analytical tools and supplementary data to visualise the findings and draw insights from the output of the frameworks. These are also steps toward verifying the outputs of GenAI to ensure that insights drawn from the framework are reliable in future applications.

The framework is applied to transcripts between 2019 and 2024, covering four stress episodes in the banking industry: (i) the outbreak of the COVID-19 pandemic in 2020H1, (ii) the beginning of Russia–Ukraine conflict in 2022H1, (iii) the US banking turmoil in 2023H1, and (iv) stresses stemming from banking exposure to commercial real estate (CRE). Figure 5 displays the average number of risk sentences extracted per transcript over time, which is higher during these stress episodes than in periods with a more stable environment.

Figure 5: Average number of risk sentences extracted from transcripts between 2019 and 2024



Source: HKMA staff estimates based on S&P Capital IQ transcript data and the S&P Capital IQ Pro database.

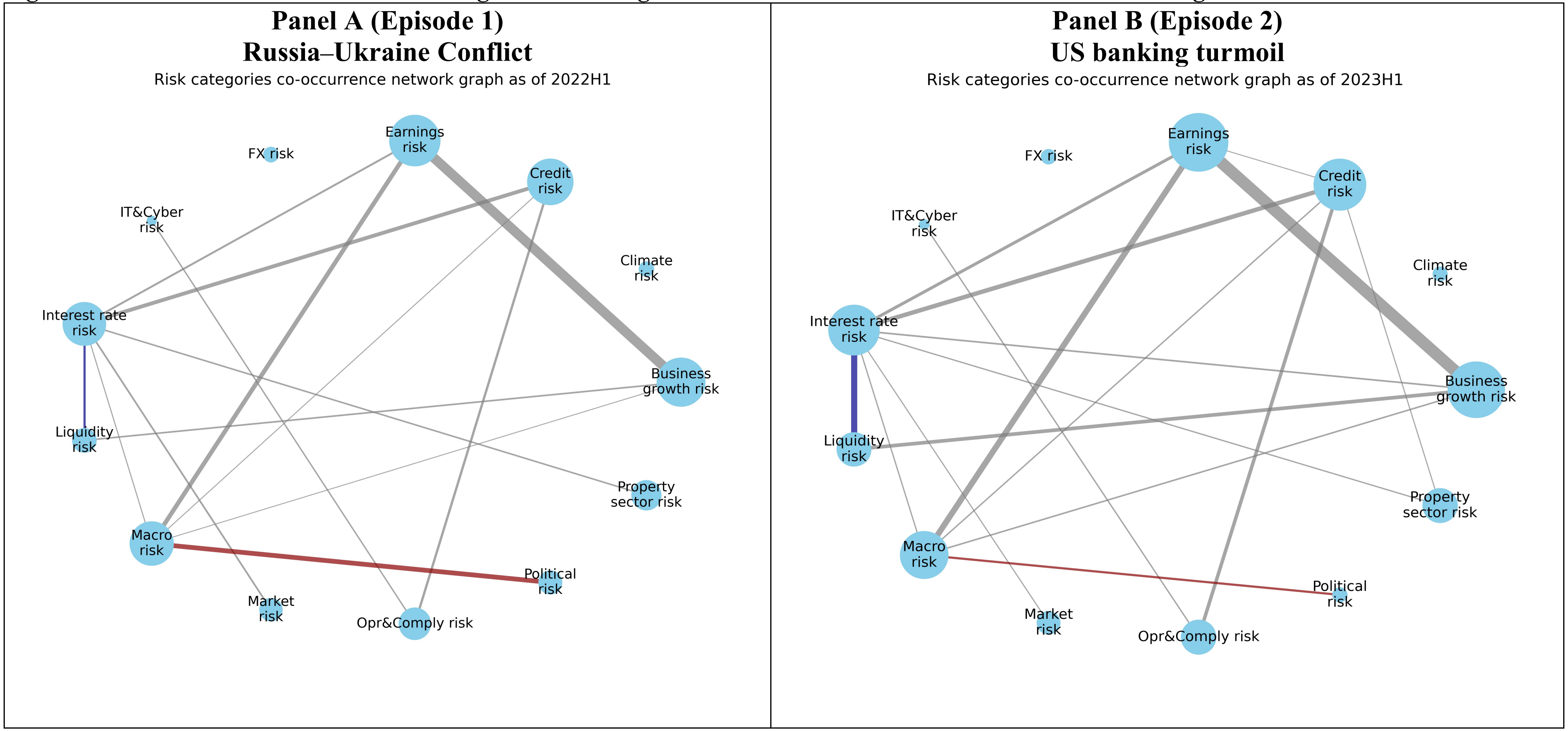
Application I: Network analysis of the Russia–Ukraine conflict and US banking turmoil

For the first application, we use a network analysis to examine the importance of and interconnections between identified risk topics during two recent stress episodes: Russia's invasion of Ukraine in 2022H1 and the US banking turmoil in 2023H1. The network analysis is particularly useful for risk surveillance because it not only highlights which risk factors have attracted increasing attention in the transcripts but also shows how different risk factors are interconnected. This can enhance our understanding of the relationships between various identified risk factors and any potential risk spillovers under different stress episodes.

We first present a network diagram of all 13 risk categories to display the underlying risk situations surrounding the two episodes (Figure 6). Subsequently, we explore the details of each episode to gain further insights.

In Figure 6, a larger node size indicates a more significant risk topic revealed from the transcripts (measured by the number of times the risk is mentioned in the transcripts relative to the corresponding level in 2021H2), and a thicker line between two risks reflects a stronger linkage between the two risks (measured by the number of times the two risks are mentioned jointly in the transcripts).

Figure 6: Relative intensities and interlinkages of risk categories as of 2022H1 and 2023H1 in the network diagrams



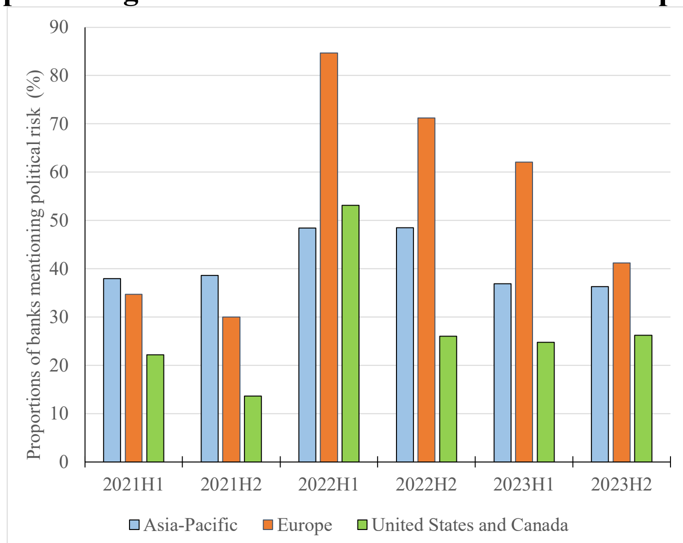
Note: A larger node size and a thicker edge width represented represent, respectively, higher frequencies (in a relative sense) of sentences mentioning specific risk categories and co-mentioning of two particular risk topics, thereby capturing their relative importance.
Source: HKMA staff estimates based on S&P Capital IQ transcript data and S&P Capital IQ Pro database.

(i) Episode 1: Russia–Ukraine conflict in 2022H1

For the first episode, the Russia–Ukraine conflict in 2022H1, as shown in Panel A of Figure 6, we find a substantial increase in the number of mentions of political risk in the transcripts (which increased 3.7 times compared with its level in 2021H2). Importantly, the increased concern about political risk is found to be closely linked with macroeconomic risks, with the number of joint mentions of both risks in the transcripts rising significantly (i.e., the red edge). These findings are consistent with banks' increased concerns about the potential negative impact of geopolitical tensions on the macro economy at that time.

Figure 7 displays the proportion of sample banks that highlight political risks in their earnings call transcripts between 2021H1 and 2023H2, segmented by three major regions. As can been seen, the rise in political risk discussion in the transcripts is greater for banks in Europe (close to 85% of banks) than banks in the US and Canada and in Asia-Pacific.

Figure 7: Proportions of banks highlighting political risk in call events by geographical regions between the 2021H1 and 2023H2 periods

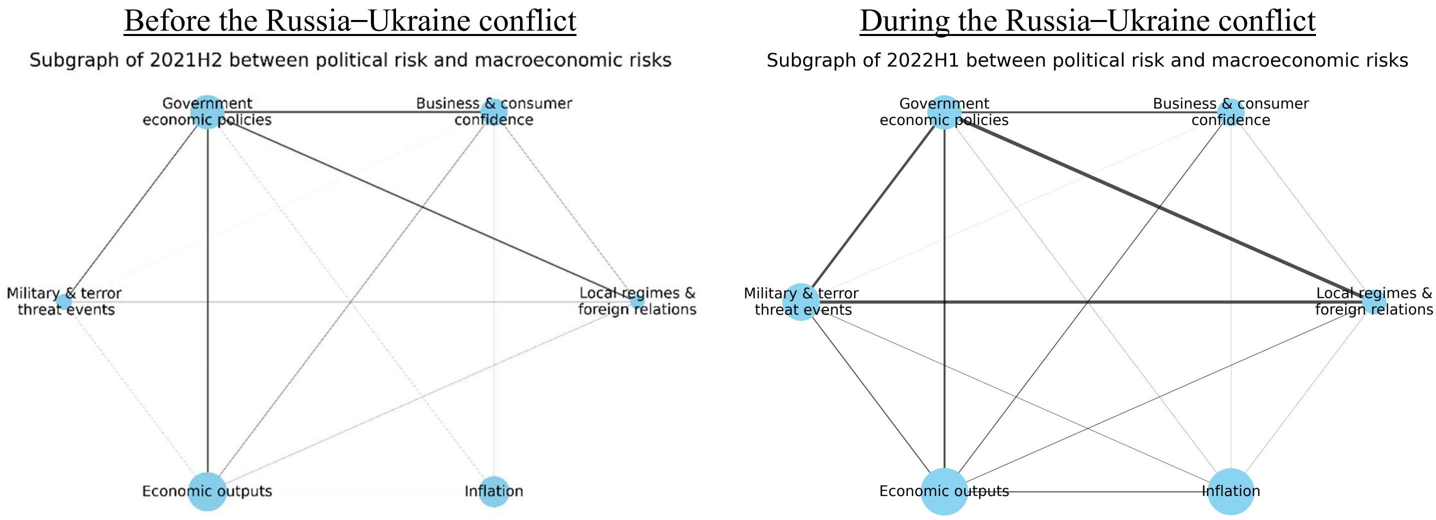


Note: The locations of the sampled banks are sourced from the S&P Capital IQ Pro database. The results for the Middle East and Africa, and for Southern and Latin America are omitted because of the limited numbers of banks in these groups.
Source: HKMA staff estimates based on S&P Capital IQ transcript data and the S&P Capital IQ Pro database.

As political risks are found to be closely linked with macro risks in the network chart (i.e., Panel A of Figure 6), Figure 8 depicts the specific network branches to examine which types of macro risk factors are most closely tied to the rising concerns about political risks. Comparison of the network branches between political risks and macro risks before (i.e., 2021H2) and during the episode (i.e., 2022H1) reveals that the rise in the number of risk discussions on "military and terror threats" is more closely linked with the "economic outputs".

Moreover, there was also a new edge between "military and terror threats" and "inflation" in the 2022H1 period, suggesting that there was rising concern in the global banking sector regarding potential economic output losses and inflationary pressure stemming from the Russia–Ukraine conflict.

Figure 8: Relative intensities and interlinkages of macroeconomic risk and political risk as of 2021H2 and 2022H1 in network diagrams



Note: A larger node size and a thicker edge width represent, respectively, higher frequencies (in a relative sense) of sentence mentioning specific risk categories and co-mentioning of two risk topics, thereby capturing their relative importance. "Military & terror threats events" and "Local regimes & foreign relations" are grouped under the category of "Political risk", and the other four risks in the chart are grouped under the category of "Macroeconomic risk".

(ii) Episode 2: US banking turmoil during the 2023H1 period

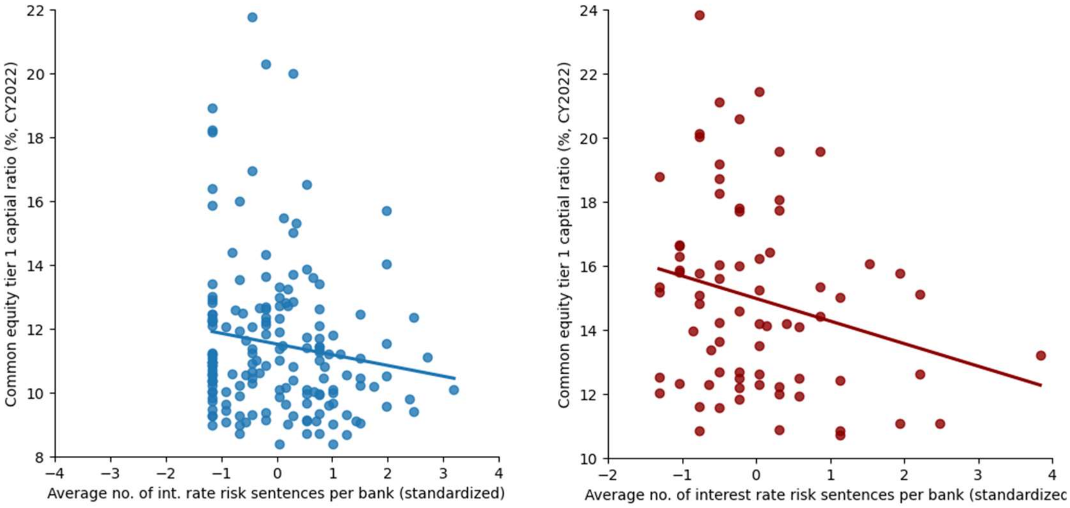
Regarding the second episode, global interest rates rose rapidly in 2023H1, and interest rate risk is identified as a key risk factor for banks in our framework (see the right panel of Figure 6). Furthermore, the increased concerns about interest rate risk are found to be strongly connected with liquidity risk, with the number of joint mentions of these two risks surging by 2.6 times (i.e., the purple edge in Panel B of Figure 6). This finding reflects heightened market concerns about banks' funding and liquidity positions following the failure of Silicon Valley Bank in the high interest rate environment.

The extent of the increase in the number of mentions related to interest rate and liquidity risks varies across banks, which may be dependent on their financial strength. We find a negative relationship between the number of mentions of interest rate risks in the Q&A section per transcript and banks' common equity tier-1 capital ratio as of calendar year 2022 (represented by blue dots) among banks in the US and Canada (i.e., the left panel of Figure 9). This suggests that banks with lower capital ratios tend to have more discussions between management and analysts on interest rate risk in the Q&A section from the transcripts data in that period. This probably reflects the fact that the US banking turmoil was triggered by an erosion of capital position from the valuation losses in banks' debt securities holdings, so less capitalised banks may have attracted more discussions on interest rate risk in the Q&A section during the events, as they had fewer buffers to absorb potential debt valuation losses from rising interest rates.

The rising concerns about interest rate risk are not confined to banks in the US and Canada, but are also evident in the transcripts of banks headquartered outside the US. Similar to the findings for US banks, we observe that better capitalised banks in other regions have fewer mentions of interest rate risks in their transcripts, as indicated by the red dots in the right panel of Figure 9. This suggests that although the US banking turmoil mainly impacted banks in the US, the incident also prompted a global reassessment of banks' interest rate risk management practices. This finding underscores the importance of maintaining a strong bank capital position for safeguarding bank resilience and investor confidence.

Figure 9: Scatter plots of bank capital ratios and the average number of mentions of interest rate risk in the Q&A section per bank transcript

a. Banks in the US and Canada **b. Banks in other regions**

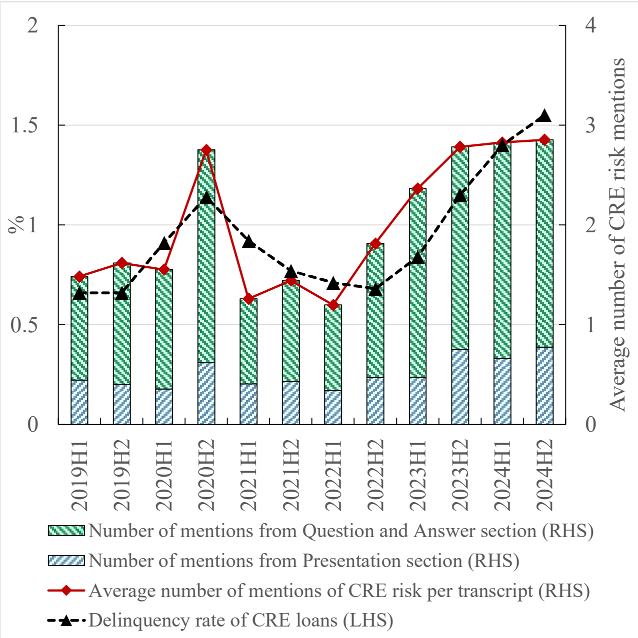


Note: In both panels, banks with call events only held between January and February 2023 (i.e., prior to the US banking turmoil in March 2023) in the 2023H1 period are excluded from the sample. In Panels A and B, we consider samples of banks headquartered in the US and Canada and in other regions, respectively, for which both interest rate risk and liquidity risk are highlighted at least once in the call events.
Source: HKMA staff estimates based on S&P Capital IQ transcript data and the S&P Capital IQ Pro database.

Application II: Early warning signal on rising risks of US-domiciled banks' exposure to the CRE market

In addition to identifying key risk factors in the transcripts, our framework is capable of providing reliable early warning signals before the stress event occurs. To illustrate this, we take the stresses stemming from US banking exposure to the CRE market driven by the rapid US monetary policy tightening as an example. Figure 10 displays the average number of CRE risks mentioned per transcripts for the sample of listed US-domiciled banks over time.

Figure 10: Average number of mentions of CRE risk per transcript for US-domiciled listed banks



Note: The second- and fourth-quarter values of the delinquency rate of CRE loans (non-seasonally adjusted) are used as the values for the first-half and second-half periods in each calendar year, respectively.
Source: HKMA staff estimates based on S&P Capital IQ transcript data and Federal Reserve Economic Data.

Two key observations are worth noting. First, amid the rapid rise in the US monetary policy interest rate since March 2022, there has been a notable rise in the frequency of CRE-related risks mentioned in the transcripts in 2022H2 (the red line in Figure 10), indicating that such risk may have emerged as a key concern. Following this, the delinquency ratio of US commercial banks' CRE loans has increased (the black dotted line in Figure 10), suggesting that the framework is capable of identifying informative signals that may indicate the potential emergence of industry-wide stress points in coming quarters.

Second, the increased CRE discussions primarily originated from the Q&A section of the transcripts. This suggests that the framework can effectively extract a wider range of information (e.g., views and assessments by analysts and investors) than that provided by banks' management, which can provide timely and valuable insights into the potential risks and vulnerabilities that banks face.

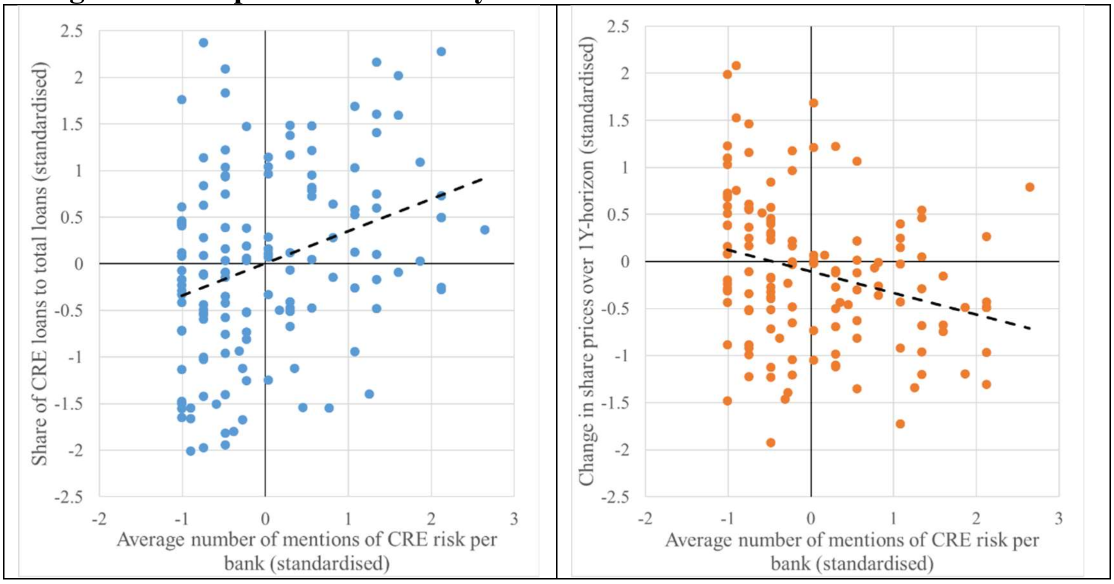
Beyond providing an industry-wide perspective, the framework can also provide useful insights for assessing CRE risk exposure at the individual bank level. The left panel of Figure 11 shows a positive correlation between the average number of mentions of CRE risk per transcript in 2022H2 and the corresponding share of CRE loans to total loans for each US-domiciled bank in 2022. This positive correlation likely reflects heightened concerns from market analysts and investors about US banks, especially those that are more exposed to CRE risk, amid a weakening CRE market in the US. This pattern also validates the usefulness of textual transcript data as a proxy for assessing banks' exposure to CRE risks.

Importantly, banks with a higher number of mentions of CRE risk in their transcript in 2022H2 tended to experience a larger drop in their equity price over the following one-year horizon, from June 2022 to June 2023 (i.e., the right panel of Figure 11). The larger decreases in bank stock prices probably reflected investors' increasingly negative perceptions of the financial resilience of these banks, which may be partly influenced by banks' CRE exposure.[14]

Together, these findings suggest that the risk indicators identified through our framework based on transcript data can provide reliable early warning signals at the industry level and can serve as useful proxies for assessing banks' CRE risk exposure at the individual bank level.

[14] Caution should be exercised in interpreting CRE exposure as the sole contributing factor to the decline in stock prices.

Figure 11: Scatter plot diagrams of the average number of CRE risk sentences extracted for each US-domiciled bank against banks' CRE loan shares and changes in stock prices over a one-year horizon



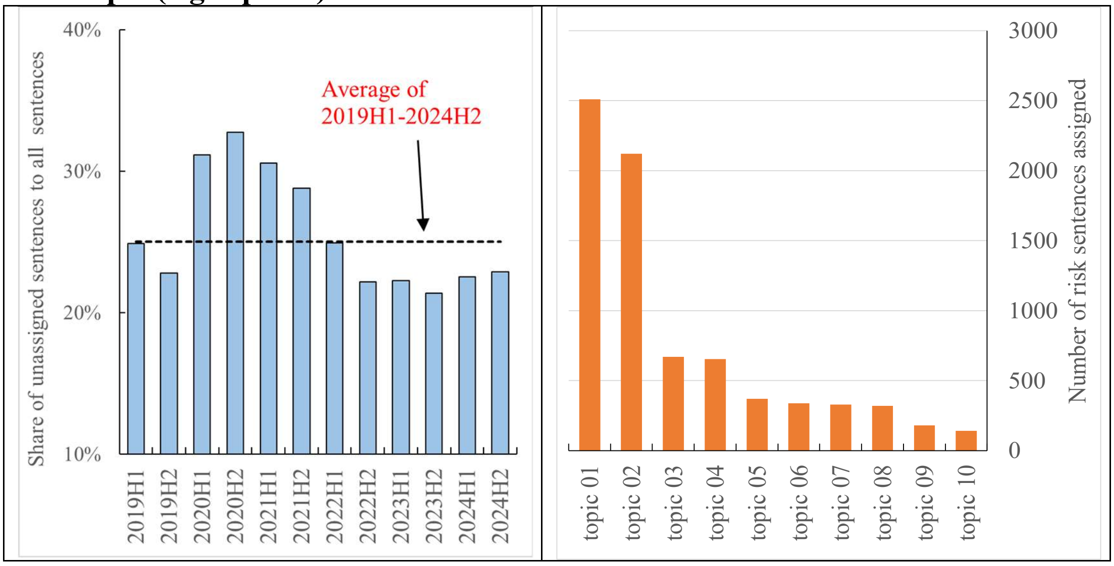
Note: In both panel, we include only banks with available data from the S&P Capital IQ Pro platform on three variables: total assets, shares of CRE loans to total loans as of calendar year 2022 and one-year horizon change in share prices between 29 June 2022 and 30 June 2023 in both panels. Observations outside of three standard deviations are omitted.
Source: HKMA staff estimates based on S&P Capital IQ transcript data and the S&P Capital IQ Pro database.

Application III: Identifying new risks from miscellaneous risk sentences using an unsupervised topic modeling technique

Next, we demonstrate how our framework is capable of extracting valuable information from the miscellaneous risk sentences to potentially identify new emerging risks and vulnerabilities. To illustrate this capability, we use the outbreak of COVID-19 in 2020H1 as an example. This period is selected for two reasons. First, as shown in the left panel of Figure 12, there is a sharp increase in the share of unassigned risk sentences extracted from the transcripts (above the average level of around 25%) during the COVID-19 pandemic episode in 2020, providing more instances for our framework to learn any new risk topics. Second, the COVID-19 pandemic was an unprecedented global shock, making it a unique case study to test whether our framework can effectively detect such a new risk factor.

As mentioned in Section III, we applied a BERTopic algorithm (an unsupervised topic model) to the group of miscellaneous risk sentences in 2020H1 to generate clusters of latent topics. Overall, more than 80 latent topic vectors were constructed, with different numbers of risk sentences assigned to them, ranging from 10 to around 2,500 sentences. We focus on the top 10 clusters (i.e., the orange bars in the right panel of Figure 12), which have at least 100 risk sentences being assigned to each of them for analysis.

Figure 12: Proportions of unassigned risk sentences between 2019H1 and 2024H2 (left panel) and number of risk sentences per latent topic assigned by BERTopic (right panel)



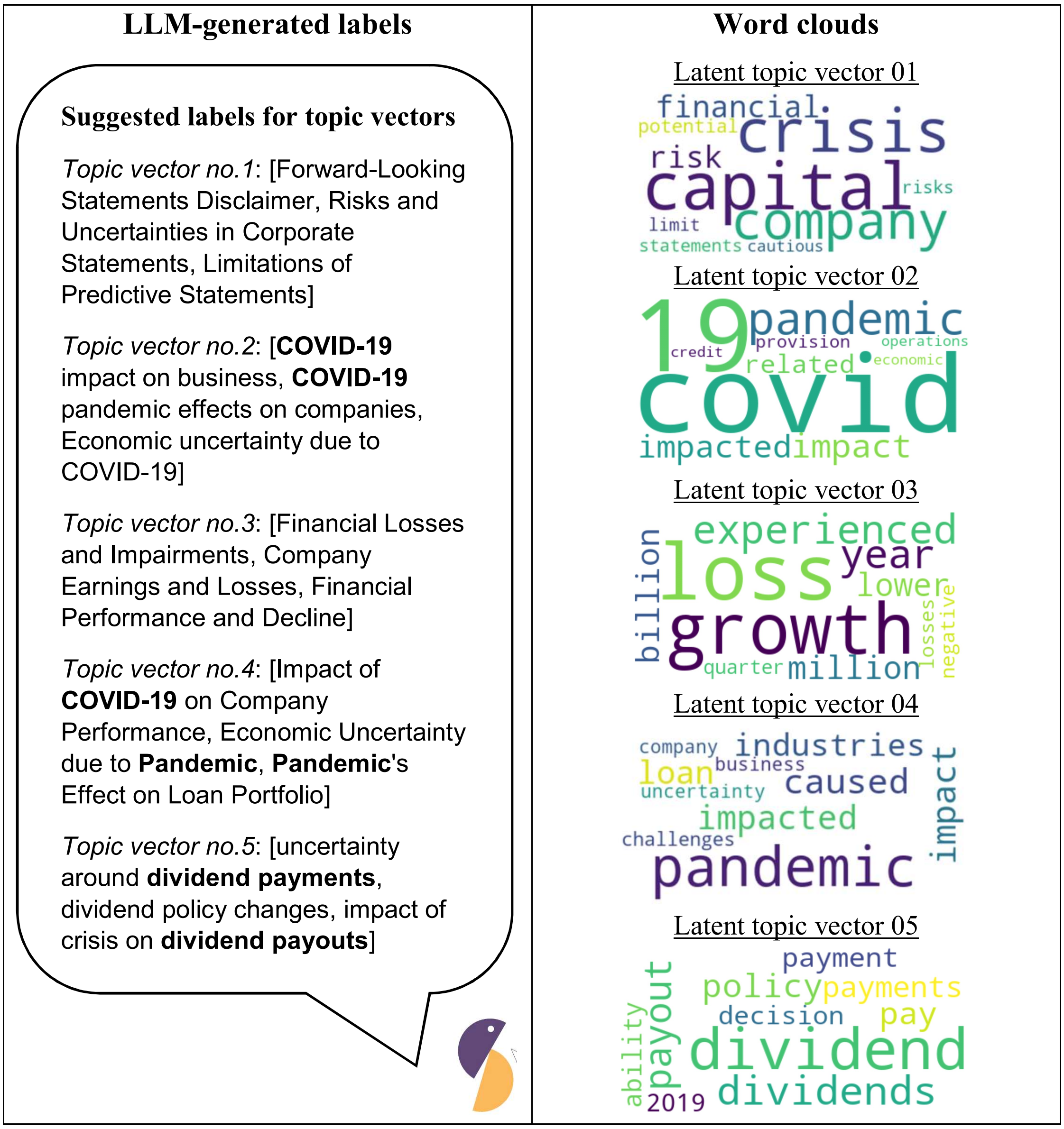
Note: In the right panel, a fraction of input risk sentences will not be assigned to any latent topic vector by the BERTopic clustering algorithm.
Source: HKMA staff estimates based on S&P Capital IQ transcript data.

For these 10 latent topic vectors, various analytics, including word clouds and LLM labeling techniques, are used to generate interpretable and meaningful topic labels. Figure 13 displays the LLM-generated label outputs and the respective word clouds for the top five latent topic vectors clustered by the BERTopic algorithm in the left and right panels, respectively.

Overall, our findings suggest that these unsupervised topic modeling techniques can effectively identify newly emerging risks at that time. Both methods can clearly demonstrate that a substantial portion of unassigned risk sentences (for instance, risk sentences from latent topic vectors 2 and 4) are related to the pandemic-driven disruptions in 2020. Uncertainty related to bank dividend payouts (i.e. latent topic vector 5) is also identified, which is in line with the challenging economic landscape and potential change in regulatory responses related to dividend policy during the pandemic[15].

[15] In response to the widespread impact of the COVID-19 pandemic in March 2020, some banking supervisory authorities (e.g. the Bank of England, the European Banking Authority, the Australian Prudential Regulation Authority) promptly issued guidance or imposed restrictions on banks' dividend distributions to shareholders (Hardy, 2021). Given the uncertainty surrounding the effects of the pandemic on economic activities, many banks were uncertain about how long these dividend payout restrictions would remain in effect.

Figure 13: LLM-generated labels and word clouds for the top 5 BERTopic clustered latent topic vectors based on 2020H1 data



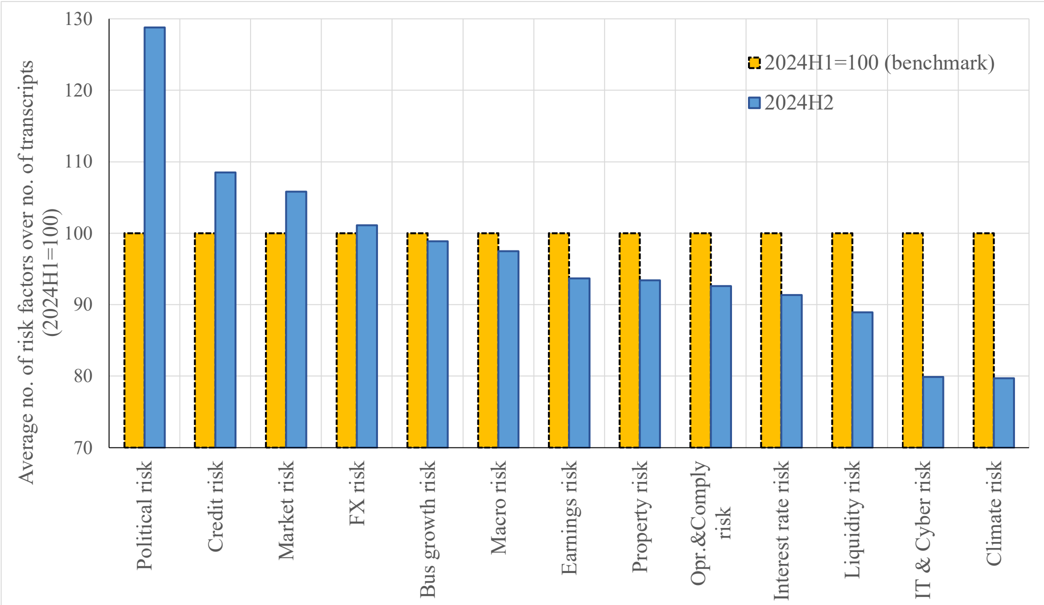
Note: For the LLM-generated labels in the left panel, labels are generated for each topic separately.

Taking together all of the findings in this section, we have shown that the framework is capable of identifying key and emerging risks facing banks globally. However, despite its usefulness for monitoring systemic risks, it is essential to recognise some limitations inherent to this new technology to provide a more balanced understanding of the potential of LLM applications in our proposed framework. These limitations are described in detail in Appendix V. With these limitations in mind, the framework should be applied in conjunction with traditional monitoring tools to mitigate the potential modeling risks.

V. LATEST ASSESSMENT BASED ON 2024H2 DATA

In this section, we apply our framework to the latest available earnings call transcripts in 2024H2 to examine the evolution of risks facing the global banking industry. [16] Figure 14 presents the changes in the average numbers of mentions for all thirteen risk categories between 2024H1 and 2024H2. The figures are all rebased to their corresponding values in 2024H1.

Figure 14: Dynamics of risk categories between 2024H2 and 2024H1



Note: Indices for all 13 risk categories in 2024H2 and 2024H1 are rebased to their corresponding values in 2024H1.
Source: HKMA staff estimates based on S&P Capital IQ transcript data.

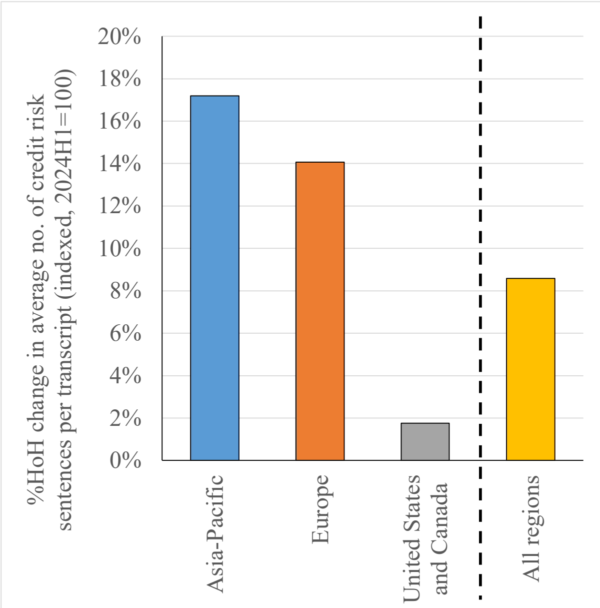
Our latest assessment shows that the global banking stability in 2024 H2 remained broadly stable and that some risk categories even saw improvement compared with the 2024H1 period (the blue bars versus the yellow bars in Figure 14). Specifically, concerns about interest rate and liquidity risks decreased notably in 2024H2, in part because several major central banks in advanced economies lowered their policy rates during the period.

However, concerns over credit risks increased. This probably reflects that despite the policy rate cuts, the possible "high-for-longer" interest rate environment may have continued to exert pressure on borrowers' debt servicing capacity. As evident in Figure 15, the rises in credit risk concerns are more attributable to banks domiciled in Asia-Pacific and Europe (the blue and orange bars, respectively), whereas the related concerns were relatively milder for banks

[16] Available transcripts between July and December 2024 are collected as of 28 December, 2024 for the assessment.

domiciled in the US and Canada. These differences may be partly attributable to a stronger-than-expected US economic environment in 2024, whereas economic growths have been relatively less positive in Asia-Pacific and Europe.

Figure 15: Increases in credit risk discussions per transcript by bank geographical locations

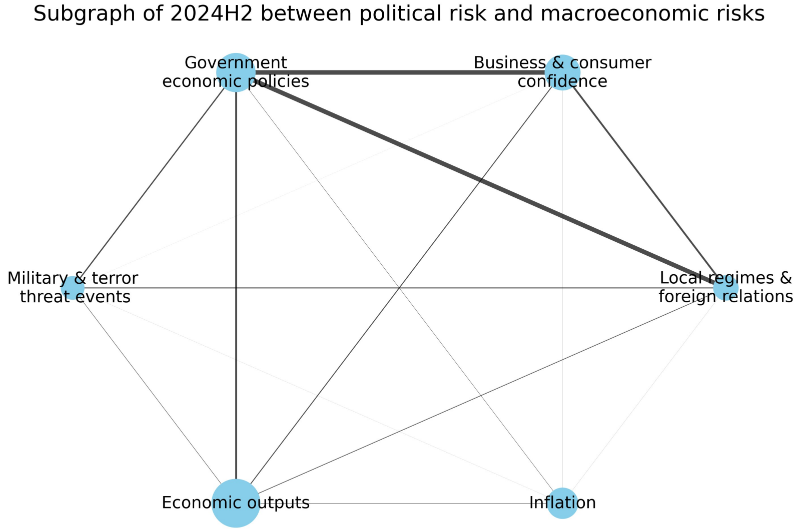


Note: The columns show the %HoH increases in the average number of credit risk sentences extracted from transcripts from 2024H1 to 2024H2.
Source: HKMA staff estimates based on S&P Capital IQ transcript data.

In addition to credit risk concern, political risks appear to be a renewed major concern, probably reflecting the heightened uncertainties surrounding US foreign policies, particularly related to trade and geopolitical issues, under the new US administration.

As shown in Figure 16, the rising concerns about political risk are found to stem from risk related to "local regimes and foreign relations", which may reflect the rising market concerns about the uncertainties surrounding economic policy proposals (especially on trade tariffs) during the 2024 U.S. presidential election. Importantly, the concerns over such economic policy uncertainty are found to be closely linked to two other macroeconomic risk factors, namely "government economic policies" and "business and consumer confidence" (as indicated by the edges in Figure 16). These findings suggest that the heightened US political uncertainties on trade policies and foreign relations, along with their negative impact on business confidence, may emerge as key concerns for the global banking sector.

Figure 16: Relative intensities and interlinkage of macroeconomic risk and political risk as of 2024H2



Note: Please see the note on Figure 8.
Source: HKMA staff estimates based on S&P Capital IQ transcript data.

It is important to emphasise that an increase in discussions about specific risks in the transcripts may reflect the perceptions of risks by bank management and market analysts rather than the actual materialisation of risks. Therefore, a rising number of risk concerns may not necessarily imply that a stress event is imminent. Nonetheless, given that political and credit risks have recently attracted increasing attention in the global banking industry, closer monitoring of these two areas may be warranted.

VI. CONCLUSIONS

This study introduces a framework that uses GenAI tools to extract useful insights from banks' earnings call transcripts for monitoring risks in the global banking sector. The framework is demonstrated to be effective in providing timely and important insights at both the industry and individual bank levels.

We apply the framework to a large sample of banks' earnings call transcripts between 2019 and 2024, and our findings show that the framework is capable of identifying key risk factors in these transcripts that align with actual risk drivers observed during recent industry-wide stress episodes, such as the Russia−Ukraine conflict and the US banking turmoil. By using various data visualisation tools, including network analysis, the framework enables us to examine the interconnections between key identified risk factors, which facilitates the monitoring of potential risk spillovers.

Moreover, the framework is found to be able to provide reliable early warning signals ahead of the actual deterioration in financial indicators prior to the stress events. Taking the stresses stemming from US banking exposures to commercial real estate (CRE) as an example, the framework is able to signal the CRE-related risk in the transcripts of US-domiciled banks ahead of the actual deterioration in the delinquency ratio of US banks' CRE loans. Together, these analyses demonstrate that the framework is useful for monitoring systemic risks in global banking.

Our latest assessment, based on transcripts available in 2024H2, reveals that credit and political risks are two important risk factors that have recently gained increasing attention in the global banking industry. Increased discussions on these two risks probably reflect continued credit quality concerns under the possible "high-for-longer" interest rate environment and renewed uncertainties surrounding trade tariffs and foreign relations under the new US administration. Thus, closer monitoring of these two areas is warranted.

Finally, although the framework has been demonstrated to be useful for monitoring systemic risks, it is important to recognise the limitations of GenAI, including its 'black box' nature. Therefore, it should be applied alongside traditional monitoring tools to mitigate potential modeling risks.

References[17]:

Andersson, M., Guillotin, J., and Neves, P., (2024). Insights from earnings calls – what can we learn from corporate risk perceptions and sentiment?, *Economic Bulletin Boxes*, European Central Bank, Vol. 4.

Cook, T. R., Kazinnik, S., Hansen, A. L., and McAdam, P. (2023). Evaluating local language models: An application to financial earnings calls. *Research Working Paper*, No. 23-12, Federal Reserve Bank of Kansas City. http://doi.org/10.18651/RWP2023-12.

Gambacorta, L., Kwon, B., Park, T., Patelli, P., and Zhu, S. (2024). *CB-LMs: Language models for central banking* (No. 1215). Bank for International Settlements.

Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., ... and Vasic, P. (2024). The llama 3 herd of models. *arXiv preprint*, arXiv-2407.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint*, *arXiv:2203.05794*.

Hardy, B. (2021). Covid-19 bank dividend payout restrictions: Effects and trade-offs. *BIS Bulletin*, No. 38, March 2021.

Hollander, T.H.S., Kalyani, A., Schwedeler, M., Tahoun, A., and van Lent, L., (2024). Economic surveillance using corporate text. *Federal Reserve Bank of St. Louis Working Paper*, No. 2024-022. URL https://doi.org/10.20955/wp.2024.022

Jha, M., Qian, J., Weber, M., and Yang, B. (2024). Harnessing generative AI for economic insights. *arXiv preprint*, *arXiv:2410.03897*.

Kim, A., Muhn, M., and Nikolaev, V. (2023). From transcripts to insights: Uncovering corporate risks using generative AI. *arXiv preprint*, *arXiv:2310.17721*.

Kwon, B., Park, T., Perez-Cruz, F., and Rungcharoenkitkul, P. (2024). Large language models: A primer for economists, *BIS Quarterly Review*, December 2024

Li, K., Mai, F., Shen, R., and Yan, X. (2021a). Measuring corporate culture using machine learning. *The Review of Financial Studies*, *34*(7), 3265-3315.

Li, K., Liu, X., Mai, F., and Zhang, T. (2021b). The role of corporate culture in bad times: Evidence from the COVID-19 pandemic. *Journal of Financial and Quantitative Analysis*, *56*(7), 2545-2583.

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., and Liang, P. (2024). Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, *12*, 157-173.

[17] Including references cited in the appendices only.

Manralai, (2023). Cosine similarity formula explained in easy words?, *Medium*, 9 March 2023. https://manralai.medium.com/cosine-similarity-formula-explained-in-easy-words-d501e42a3cb5 on 20 February 2025.

Masson, C. (2023): Automated analysis of risk factors published by listed companies: A use case of natural language processing for the AMF. *Report/Study*, Autorité des marchés financiers (AMF), January 2023.

Reimers, N., and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *arXiv preprint*, *arXiv:1908.10084*.

Sarkar, S. K., and Vafa, K. (2024). Lookahead bias in pretrained language models. Available at SSRN: http://dx.doi.org/10.2139/ssrn.4754678.

Soto, P. (2019). Breaking the word bank: Measurement and effects of bank level uncertainty, *Journal of Financial Services Research,* 59 (1–2), 1–45.

Wang, M., Chen, L., Cheng, F., Liao, S., Zhang, X., Wu, B., ... and Li, Y. (2024). Leave no document behind: Benchmarking long-context LLMs with extended multi-doc QA. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 5627-5646).

Wei, L., Li, G., Zhu, X., and Li, J. (2019). Discovering bank risk factors from financial statements based on a new semi-supervised text mining algorithm. *Accounting & Finance*, 59(3), 1519-1552.

Appendices

I. *Additional information about the Llama 3 model used in this study*

The Llama 3 models developed by Meta Inc. are a series of open-source large language models (LLMs) released in 2024. These models are designed with enhanced natural language understanding capabilities to be suitable for various text generating applications across a wide range of topics. They were pre-trained on around 15 trillion tokens of data from publicly available sources, including publicly available instruction datasets, as well as over 25 million synthetically generated examples.[18] According to information[19] provided by Meta, the knowledge cut-off date of the Llama 3 series models from the publicly available online training data is December 2023. The models achieved good performance on a number of benchmarks and have been widely applied in economic and finance studies because of their transparency and open-source nature. Our text summarisation results based on the LlaMa 3-70B model are based on the parameters listed in Table A.1.

Table A.1: Selected hyper-parameters for the LLM summarization task

Hyper-parameters	Value
Temperature	0.05
Top_p	0.9
Maximum no. of output tokens	512
Repetition penalty	1

II. *Chunking and text pre-processing*

We follow the literature in chunking the contents of the transcripts into several sub-sections for LLM processing,[20] for two reasons. First, earlier works indicate that chunking can alleviate the issue of deteriorating quality in LLMs generated outputs when LLMs are processing very long context documents (i.e. those with at least 8,000 tokens). Second, the length of the context windows that can be processed by LLMs will be limited by the capacity of the underlying graphics processing unit hardware.

To strike a balance between providing sufficient contents for textual understanding and maintaining a manageable total length of the given inputs for LLM processing, we apply a maximum threshold of 2,000 input tokens (based on the nltk package) for chunking transcripts into sub sections, following the approach outlined in Kim, Muhn and Nikolaev (2024). In line with their study, we separate the presentation and Q&A sections before any chunking steps. We avoid splitting consecutive paragraphs of the same speech by same executive in the presentation unless the total length of the consecutive speech exceeds the 2000 token thresholds. For the Q&A sections, we consolidate individual questions by analysts and corresponding answers by executives into the same chunked section. In addition, the names and role titles of

[18] Relative to other closed-source pre-trained LLMs, the specific details of the training data and model structures for Llama3 series models are more transparent.
[19] https://docs.api.nvidia.com/nim/reference/meta-llama-3_1-405b.
[20] While there have been recent advancements in extending context windows that can be processed by a LLM, their performance remains to be examined as Liu et al. (2024) and Wang et al. (2024) show that the performance of LLMs on challenging tasks decreased considerably when the context length increased.

bank management and analysts are deleted from the inputs as this information is assumed to be irrelevant in the summarisation task. After completing the chunking step, we will use application programming interface to feed all of the chunked sections of transcripts together with the designed prompt through the LLM iteratively for processing.

III. Fine-tuning the word embedding model in topic classification

We conduct additional fine-tuning on the sentence transformer model to enhance its capacity to identify banking-specific terminology. One shortcoming that may arise when applying pre-trained word embedding models in the finance domain is that they are trained on general text and therefore may not have sufficient understanding of banking industry terminology. For example, Gambacorta et al. (2024) discussed the extent to which existing pre-trained models have sufficient understanding of central banking terminology and recommended fine-tuning to improve model performance. To address this concern, we conduct additional fine-tuning on the pre-trained embedding model by providing more than 300 labeled risk discussion sentences corresponding to the 37 topics for further learning. These labeled risk sentences are mainly sourced from risk sentences generated by LLM on bank transcripts between 2016 and 2018, with some additional sentences handcrafted to balance the number of labeled examples across the pre-defined 37 risk topics. Selected labeled examples are listed in Table A.2 for illustration.

For the fine-tuning technique, we adopt the contrastive learning approach with triplet loss function to refine the "bge-large-en-v1.5" embedding model with the labeled risk sentences. A fine-tuned embedding model is used for deployment after three epochs as a larger number of epochs could be prone to overfitting.

Table A.2: Selected examples of risk sentences and their risk topic labels

Labelled risk sentences	Risk topic assigned
Non-performing loans (NPLs) were up a little bit in the retail space.	Credit risk (Non-performing assets)
The provisioning outlook for 2018 is challenging, with a cost of risk annualized at 70-80 basis points, despite expected growth of 10-15%.	Credit risk (Loan loss provisions & charges)
Foreign currency risk-weighted assets may increase due to RMB depreciation.	Foreign exchange risk (Abrupt movements in currency exchange rates)
The asset is mismatched due to the short duration of the retail book, which is mostly fixed rate, and the re-pricing of fixed deposits on maturity.	Interest rate risk (Monetary policy and funding costs)
Deteriorating macro prospects in Brazil, with some downgrades in consensus forecast.	Macroeconomic risk (Economic output)
The company has had to strengthen its anti-money laundering (AML) capabilities, which may indicate past weaknesses in this area.	Operational & compliance risk (Anti-money laundering)
The speaker notes that there are legitimate complaints about trade, implying that there are underlying issues with the current trade policies.	Political risk (Local regimes and foreign relations)

IV. Implementation of BERTopic for unsupervised topic modelling

In deployment, we largely follow the setting recommended by the BERTopic architecture's author, which is available on the model webpage.[21] Some key model settings and hyper-parameters are reported in Table A.3 below. Nevertheless, additional calibration of hyper-parameters to optimize the model performance may be warranted given that the number of unassigned risk sentences could vary in each update.

Table A.3: Selected hyper-parameters for BERTopic implementation

Hyper-parameters	Value
Model version	0.16.4
Pre-trained general purpose word embedding model	bge-large-en-v1.5
Bigram representation detection	Yes
Removing English Stop-words	Yes
Minimum frequency of word occurrence	5
Minimum topic size	10

For the topic visualisation techniques, word cloud technique is already embedded in the BERTopic architecture. Regarding the LLM-labelling approach, we can provide 20 representative sentences for each of the first 20 latent topic vectors to the LLM (i.e., the Llama3-70B model mentioned above) to generate human-interpretable labels. The prompt template for instructing LLM to generate topic labels (crafted by the GPT-4o model) is shown below in Figure A.1.

Figure A.1: Prompt template for instructing LLM to generate topic labels for latent topic vectors from BERTopic

Prompt template
System: You are a helpful, respectful and honest assistant for labeling topics. You follow instructions extremely well. Return answers after ####. I have a topic that contains the following documents: 1. Traditional diets in most cultures were primarily plant-based with a little meat on top, but with the rise of industrial style meat production and factory farming, meat has become a staple food. 2. Meat, but especially beef, is the word food in terms of emissions. 3. Eating meat doesn't make you a bad person, not eating meat doesn't make you a good one. The topic is described by the following keywords: 'meat, beef, eat, eating, emissions, steak, food, health, processed, chicken'. Based on the information about the topic above, please create a short label of this topic. Make sure you to only return the label and nothing more. #### environmental impacts of eating meat I have a topic that contains the following documents: {INSERT 20 representative risk sentences for each latent clustered topic vector} Based on the information about the topic above, please create at most three short labels of this topic. Make sure you to only return the labels and nothing more after ####.

[21] https://maartengr.github.io/BERTopic/index.html

V. Limitations of applying LLM technology to transcript data

While the previous analyses demonstrate the benefits and capabilities of LLM technology, it is essential to recognise some limitations inherent to this new technology to provide a more balanced understanding of its potential in our proposed framework.

First, one major challenge in the application of LLMs is related to their "black-box" nature in text generation, as they are prone to generating untruthful contents because of their stochastic nature. They are also susceptible to the "look-ahead bias" problem, which may occur if the training dataset of an LLM includes information that theoretically should not be known to earnings call event participants during the event, such as the financial and operating performance of a bank after the call event. In practice, it is challenging to eliminate this issue because of the vast amount of training data sourced from the Internet to train the latest LLMs and the complexity of LLM architecture. In Appendix VI, we discuss theoretically how our framework might also be subject to the look-ahead bias, and we will conduct some simple exercises to examine whether the look-ahead bias is a major concern in our framework. In short, we do not find strong evidence to suggest the existence of the look-ahead bias problem in our framework.

Another limitation of our current analytical framework is that it may not be adaptive to future behavioral and linguistic changes in the discussions between bank management and analysts during call events. In response to the wider usage of AI text processing applications, bank management and analysts may strategically reshape the contents and word choices in their textual disclosures to receive more favorable assessments from AI text processing tools. Such behavioral changes could weaken pre-trained LLMs' capacity in analysing and summarising risk content from the transcript data if they have only been trained on textual data with limited representations of these strategically reshaped text disclosures.

In principle, these two limitations can be partially addressed if an open-source pre-trained LLM is regularly updated and an exact knowledge cut-off date for the training or the fine-tuning dataset is available for application. When a specific LLM is only applied to transcript data that are available only after the knowledge cut-off date, then its usage will not be subject to concerns about look-ahead bias. Nevertheless, given the intensive resources required to train a high-performance LLM at this time, achieving these conditions may be challenging for small and medium-sized institutions outside the technology sector.

VI. Assessing the look-ahead bias concern in outputs generated by LLMs

In our use case, the existence of look-ahead bias could imply that the LLM may have acquired knowledge about the "future" outcomes of a bank from various other sources in the training dataset that were only available after call events. These sources could include news articles, announcements from banks and regulators, market intelligence reports, and other forms of data from the Internet. In the summarisation task, by identifying the particular bank associated with the provided transcript section, the LLM could also retrieve other relevant information from the abovementioned sources that is embedded within the parameters of the LLM to complete the summarisation task. This would deviate from the original intention of our use case, which is to use only the given transcript section as input to summarise and generate bullet point risk sentences. Although several techniques have been proposed to alleviate this problem, the effectiveness of these techniques in empirical applications remains an open research question.

As a robustness check, we follow the spirit of the approach adopted by Sarkar and Vafa (2024) to test whether the LLM generated any risk sentences directly referencing the “pandemic outbreak” during the 2019 period. The COVID-19 pandemic was an unanticipated event that is not supposed to have been predictable by any market participants. If we found references to it among risk sentences in the LLM generated output for the 2019 period, this would be a strong indication of look-ahead bias in LLM applications. After screening all of the risk sentences from 2019, we find no risk sentences referencing terms related to the pandemic era (e.g., “Covid”, “pandemic”, “respiratory disease”, etc.), suggesting that LLM drew mainly on the given input for summarisation.

We also check whether LLM accessed future data related to the Russia-Ukraine conflict by examining the political risk discussion contents in the 2021H2 period. A single-digit number of risk sentences mentioned changes in the political landscape in Ukraine. These contents were drawn from the underlying discussions in call event during the same period and was based on the escalated border tensions between Russia and Ukraine at the end of 2021. No risk sentences generated by the LLM in the entire 2021 period included wording that directly reflected the military invasion (e.g., “war”, “military”, “invasion”, etc.).

Our proposed framework also incorporates other analytical processing steps on LLM-generated outputs to further scale down our reliance on directly inferring outputs generated by the LLM only. Taking all of these findings together, we consider the look-ahead bias problem associated with LLM technology to be controllable in this study.

VII. Additional Figures and Tables

Figure A.2: Illustrative example of an original excerpt from a transcript and the associated LLM-generated risk summary

An excerpt from the presentation section of transcript with call title "Bank of Queensland Limited, 2022 Earnings Call, Oct 12, 2022" held on 11th October, 2022
"*Thank you, Racheal. In summary, we are delivering quality, sustainable profitable growth through the disciplined execution of our strategy. We've made good progress on our transformation to a truly end-to- end multi-brand cloud digital bank. The lessons learned so far make it clear that we are on the right track. Our digital transformation and simplification will bring greater benefits for our customers, and that means growth of our business into the future will be at a lower unit cost. And we will have even greater flexibility to really go after the opportunities that will deliver for our customers and shareholders. Finally, to our outlook on Slide 43. Australia remains well placed given low unemployment, high levels of accumulated household savings and high terms of trade. Consumer spending is at levels above pre- COVID, and businesses are investing as the economy recovers. However,* ***uncertainty remains given elevated inflation, rising interest rates, global tensions and slowdown and ongoing impacts to supply and labor.*** *We remain committed to delivering sustainable profitable growth.* ***We expect credit growth to slow in FY '23****. And while we'll continue to grow our market share, we see this period as a time for optimizing margins, revenue and returns. We will have tailwinds in revenue given the steady quality growth being delivered across Retail and Business Banking.* ***We have positive NIM momentum leading into FY '23 and with further tailwinds expected from rising interest rates, partially offset by the headwinds of rising funding costs.*** *We are heading into a period of* ***cost headwinds given the inflation, regulation and the dual costs of the old and new banking platforms,*** *and we are managing to positive jaws. CET1 is expected to broadly remain above 9.5%. We absolutely understand the importance of dividends for our shareholders. And our payout target range remains at 60% to 75% of full year cash earnings. Thank you very much for your time this morning. I'll now hand back to Cherie and open it up for questions.*"
LLM-generated summary of risk sentences in bullet points format based on the prompt template in Figure 3
Negative factors: 1. Uncertainty remains due to elevated inflation. 2. Rising interest rates pose a risk. 3. Global tensions and slowdown are potential threats. 4. Ongoing impacts to supply and labor are concerns. 5. Credit growth is expected to slow in FY '23. 6. Cost headwinds are anticipated due to inflation, regulation, and dual banking platforms. 7. Rising funding costs may partially offset positive NIM momentum.

Source: S&P Capital IQ transcript data and LLM-generated outputs.

Table A.4: Risk topics

No.	Risk category	Concept	Anchor sentence
1	Interest Rate Risk	Monetary policy and funding costs	Changes in monetary policy interest rates or market interest rates have affected funding costs, affecting the maturity and duration of bank's loan and deposit positions
2	Interest Rate Risk	Interest rate environment	The high interest rate environment and interest rate outlook was driven by monetary policy decisions. Interest rate hikes will create losses in fixed income portfolio.
3	Market Risk	Capital market volatility	Volatility in capital markets and financial markets about stocks, equity securities, debts securities, commodities and derivative instruments.
4	Market Risk	Investment valuation losses	Investment securities losses driven by changes in stock prices, bond valuation, fixed incomes securities prices, commodity prices, or market volatility.
5	Foreign Exchange Risk	Abrupt movements in currency exchange rates	Adverse movements in currency exchange rates such as sharp depreciation and appreciation in the currency exchange rate, such as with the US dollar, Euros, British Pounds or Japanese Yen
6	Foreign Exchange Risk	exchange rate policy regime	Currency exchange rate fluctuations driven by a change in the exchange rate policy regime of the underlying currency
7	Credit Risk	Non-performing assets	Non-performing loans (NPLs) and credit losses from loan payment overdue and higher borrower default risks and delinquency assets
8	Credit Risk	Loan loss provisions & charges	Higher loan loss provision and credit provisioning coverage, resulting in larger credit losses or credit impairment charges and impairment losses, the bank need to set aside more loan loss provision
9	Credit Risk	Weak asset & credit quality	Deterioration in asset quality of the loan assets and higher credit risks from borrowers, lower credit quality of the loan book
10	Credit Risk	Lending standard & Cost of credits	Lending standard, credit standard and underwriting standard tighten and a higher credit risk premium.
11	Liquidity Risk	deposits outflows and shortage in cash flow positions	Liquidity concerns in cash shortage and fast deposits outflows make it harder for banks to manage cash flow positions and convertible assets holdings.
12	Liquidity Risk	Unstable interbank and wholesale funding	Funding redemptions and liquidity squeeze could occur when interbank funding market to obtain new wholesale funding is not available.
13	Liquidity Risk	High loan-to-deposit ratios	High loan-to-deposit ratio or low deposit-to-loan ratio could limit bank credit supply.
14	Earnings Risk	Lower earnings and revenues	Declining profits and earnings due to thinner deposits interest margins, lower net interest income NII, lower commissions and fee incomes and lower revenue
15	Earnings Risk	higher operating costs	Higher operating costs and other expenses, higher salaries, larger manpower expenditures expenses and other operating inefficiencies in the business on cost-to-income ratio
16	Earnings Risk	reduction in profit ratios	Decrease in profitability ratios such as earnings before interest and taxes (EBIT, EBITDA), returns on asset (ROA) and returns on equity ratios (ROE)
17	Business growth Risk	Credit and deposit changes	Negative loan growth and deposit growth from intense competitions between banks, and the difficulty in retaining customers and clients in business
18	Business growth Risk	non-interest revenues	Negative growth in fees and commission business such wealth management products, credit cards services, treasury and trade services
19	Business growth Risk	Business plans & targets	Business plans, management targets and growing opportunity for new business segments have not yet achieve good results.

No.	Risk category	Concept	Anchor sentence
20	Business growth Risk	Expansions, merger & acquisition	Change in business models has limited the opportunity for business expansions and merger and acquisitions.
21	Climate Risk	Climate transition risk	Carbon emissions, greenhouse gases emissions, fossil fuels, climate change issues and sustainability.
22	Climate Risk	Climate physical risk	Climate physical risks and damages from higher occurrence of natural disasters, weather, sea levels and temperature increases
23	Climate Risk	sustainability requirement	Climate change issues, sustainability and ESG factors imposed by supervisors
24	IT and Cyber Risk	Cyber security	Cyber security issues and/or failures in information technology systems due to cyber-attacks or malware or ransomware or computer system malfunctioning
25	IT and Cyber Risk	data & internet related crimes	Internet crime events such as data leakages, data breaching, unauthorized fund transfers and digital frauds
26	Operational & Compliance Risk	Inadequate control	Inadequate internal processes, people, risk management control processes or external events results in financial frauds, regulatory compliance failures
27	Operational & Compliance Risk	Legal enforcement and litigation	Legal enforcement risks, disciplinary actions and compliance issues taken by financial regulators and supervisors
28	Operational & Compliance Risk	Regulatory requirements	Regulatory requirements, regulatory approvals and regulatory clarifications from financial industry supervisors
29	Operational & Compliance Risk	Anti-money laundering	Anti-money laundering (AML) and counter terrorist financing (CTF), customer due diligence (CDD) and know your customer (KYC) requirement burdens
30	Political Risk	Military conflicts and terror threats	Political events such as war, wars, wars threats, military conflicts, military attacks, terrorist attacks, protests, social unrests
31	Political Risk	local regimes and foreign relations	Events such as general elections, trade disputes, political and diplomatic sanctions, worsening diplomatic relations, tensions in relations between countries
32	Macroeconomic Risk	Economic output	Macroeconomic developments such as negative GDP growth, rising unemployment, economic recession, slowdown in production outputs, weak consumption
33	Macroeconomic Risk	inflation and price level	Elevated inflation rate and rapid wage increases, inflationary pressure, worry about stagflation has increased price levels of every day goods and services. Concerns about deflation and deflationary pressure
34	Macroeconomic Risk	business and consumer confidence	Weak business confidence, declines in consumptions and in business capital expenditure and business investments
35	Macroeconomic Risk	government economic policies	Unanticipated implementation of government economic policies such as new taxes, tariffs, subsidies, prohibitions on selling
36	Property sector Risk	Residential real estate, housing prices	Residential real estate and residential properties and housing prices and land prices affecting property loans and mortgage business
37	Property sector Risk	Commercial real estate	Commercial real estate and office building properties downturns leads to decline in office rentals and commercial property collaterals