**BoE Project Assignment 3: Final report**

Team STOMPA

30th June 2025

**1 Background and Context**

The Bank of England’s Prudential Regulation Authority (PRA) is responsible for ensuring the safety and soundness of critical financial institutions, including global systemically important banks (G-SIBs). While these institutions regularly publish quantitative financial data, a significant amount of qualitative insight is conveyed during earnings calls and analyst Q&A sessions, with transcripts often containing nuanced discussions of strategy, operations, and market concerns. However, their unstructured and technical nature limits use of traditional analytical methods.

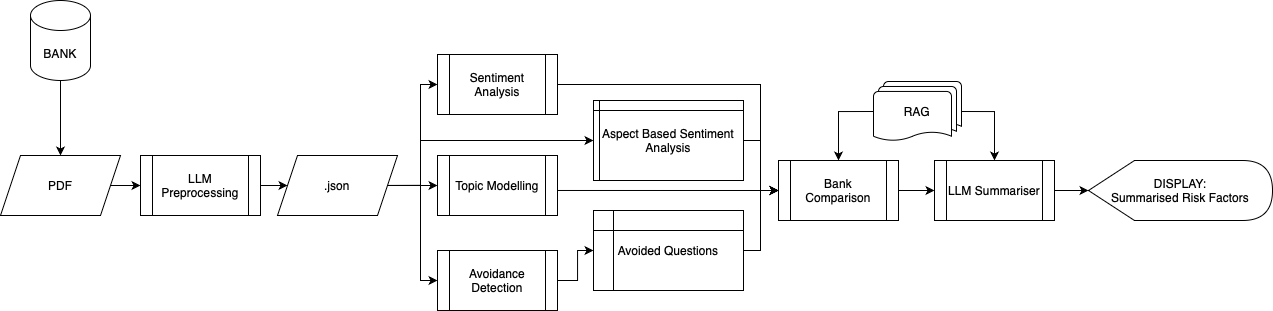
This project seeks to apply advanced natural language processing (NLP) techniques to extract and interpret relevant information from these transcripts. The goal is to enhance the PRA’s ability to detect early signals of risk and institutional change that may not be apparent in standard financial disclosures.

The central challenge is to determine whether these transcripts contain early warning signals that can inform supervisory judgment and decision-making. Specifically, we aim to identify:

* **Financial or operational risks** not evident in headline metrics
* **Strategic shifts** that may indicate changes in risk appetite or business model
* **Market concerns** raised by analysts that may point to reputational or systemic issues

Transforming unstructured transcripts into structured insights allows the PRA to strengthen its oversight and improve responses to emerging risks.

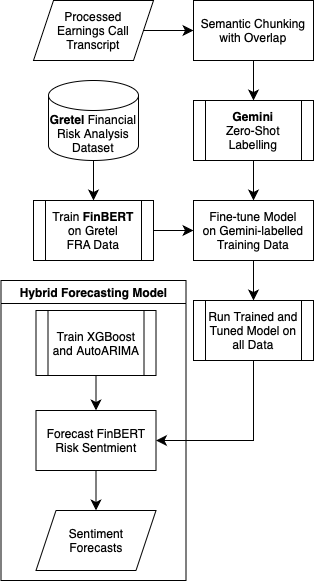
**2 Project Development Process**

The process flow shows the full pipeline for the analysis of the quarterly earnings calls. It allows for multiple agents simultaneously to analyse the data for indicators of risk and will display the key insights and findings through graphs and summaries comparing banks, quarters, sentiments and topics.



**2.1 Sentiment Analysis**

2.1.1 Risk-Based Sentiment Analysis

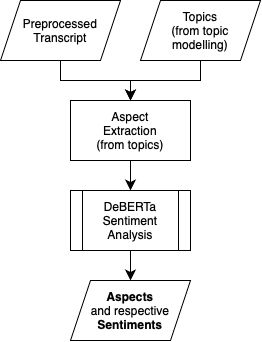
The sentiment analysis component of this project followed a structured, iterative process encompassing data acquisition, preprocessing, modelling, and evaluation. Quarterly earnings call transcripts were collected for three major G-SIBs: Citi, JPMorgan and UBS, spanning 2010 to 2022. These were processed using PyMuPDF and segmented into Presentation and Q&A sections for more granular analysis.

Risk severity labelling employed a three-stage strategy. The initial phase used a rule-based heuristic labeller, applying lemmatised unigrams, bigrams, and trigrams matched against curated keyword sets to classify sentences as LOW, MEDIUM, or HIGH risk. To improve semantic flexibility and scalability, Gemini was then used for zero-shot classification, enabling context-aware annotation at scale. These outputs, combined with the GretelAI Financial Risk Analysis dataset, were used to fine-tune FinBERT, producing a domain-specific model for consistent and automated risk tagging during inference.

For forecasting, a hybrid modelling approach was implemented by combining SARIMA and XGBoost. SARIMA parameters were selected via auto\_arima. XGBoost was trained on lagged features with early stopping. Outputs from both models were integrated using a weighted average.

Final forecasts were visualised in a Streamlit-based dashboard aligned to future quarter-end dates. Key challenges included inconsistent transcript formats and initial FinBERT misclassification of MEDIUM risk, both mitigated through refined preprocessing scripts and data augmentation.

2.1.2 Aspect-Based Sentiment Analysis

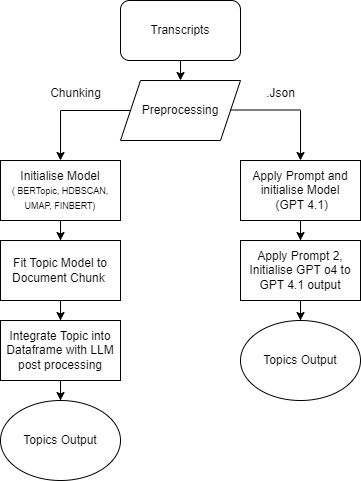
Aspect-based sentiment analysis (ABSA) was used to gain insight into the sentiment expressed by G-SIBs towards topics identified in the earnings call transcripts. This enabled nuanced interpretation of how institutions communicated around key financial concepts such as liquidity and interest rates with the aim of unveiling shifts in tone that may reflect emerging risks or changing strategic priorities.

A DeBERTa model was applied for this task (*deberta-v3-base-absa-v1.1*). This is a transformer-based model that has been fine tuned specifically for ABSA and was largely chosen for its ability to capture subtle contextual cues. This ability is important for analysing the formal language used in financial presentations and publishings.

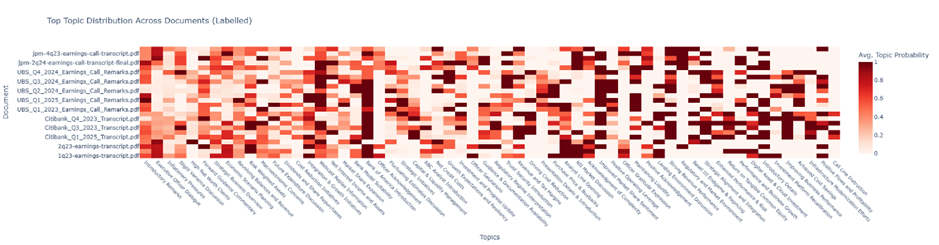
The aspects to be evaluated were derived directly from the keywords identified through the topic modelling stage of the pipeline described in the next section. This integration aims to enable consistent tracking of how each G-SIB expresses sentiment towards their most frequently discussed themes across quarters.

**2.2 Topic Modelling**

Two complementary methodologies were employed for topic extraction from banking earnings calls: BERTopic and LLM-based topic modelling. The analysis incorporated transcripts from three major banks, with Silicon Valley Bank serving as a baseline for identifying risk-associated topics in failing institutions.



BERTopic utilised UMAP, HDBSCAN, and FinBERT for financial-context embeddings, requiring custom preprocessing of raw transcripts. However, several limitations emerged: spaCy misclassified proper names, n-grams could be generic, and topic interpretability remained challenging. To address some of these constraints, a Gemini LLM post-processing step was integrated to generate contextualized topic summaries with frequency analysis.



The LLM approach processed JSON files through OpenAI's GPT-4.1 API using carefully designed prompts, with results refined through GPT-4o-mini due to enhanced token capacity. This method identified the five leading topics per transcript, keywords and summarisations. Providing granular, actionable insights compared to BERTopic's broad categorizations.

Key advantages of the LLM methodology included highly specific thematic clustering, problem-focused identification of challenges e.g Net Interest Margin pressure and Commercial Real Estate exposure. This context and justification enhances the RAG's capacity to deliver accurate, relevant responses.

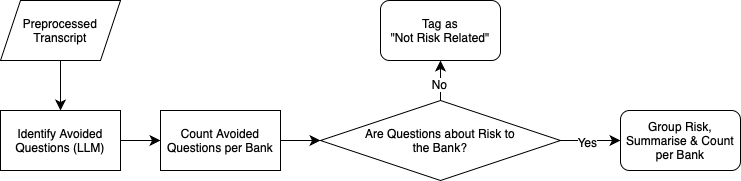
Validation against Silicon Valley Bank suggests the approach's effectiveness in identifying high-risk positions, though temporal context limitations are acknowledged.

2.2.1 Future Development

Cost optimization is essential as LLM processing expenses escalate with scale, requiring exploration of hybrid approaches and quantitative performance metrics to objectively compare methodologies. Expanding validation beyond Silicon Valley Bank to include multiple distressed institutions would enable statistical analysis of correlations between topic patterns and actual bank risk levels.

**2.3 Question Avoidance Detection**

This section assesses whether executives evaded direct answers to analyst questions with the aim of revealing notable trends in transparency and executive communication strategy. This process is based on Google’s gemini-2.0-flash LLM which analyses transcripts and highlights any avoided questions and its relevant theme.

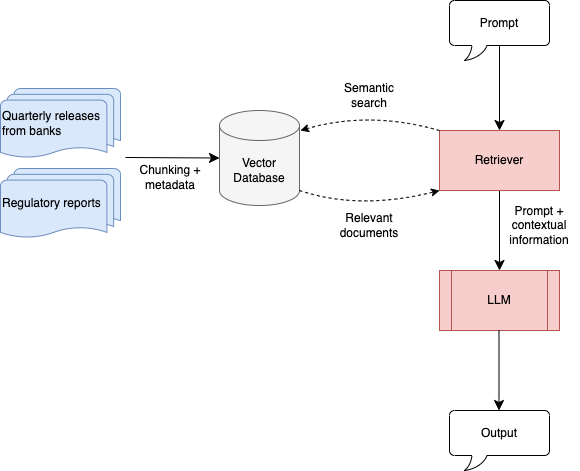
For each Q&A pair, the model evaluates whether an answer can be considered *evasive*. The prompt is engineered to classify responses with clarity and objectivity, marking them as either “avoided” or “answered.” Avoided questions are then noted and stored for analysis of any overarching themes.

**2.4 Retrieval Augmented Generation (RAG)**

LangChain has been utilised to develop a RAG pipeline. A vector database has been populated using two types of information:

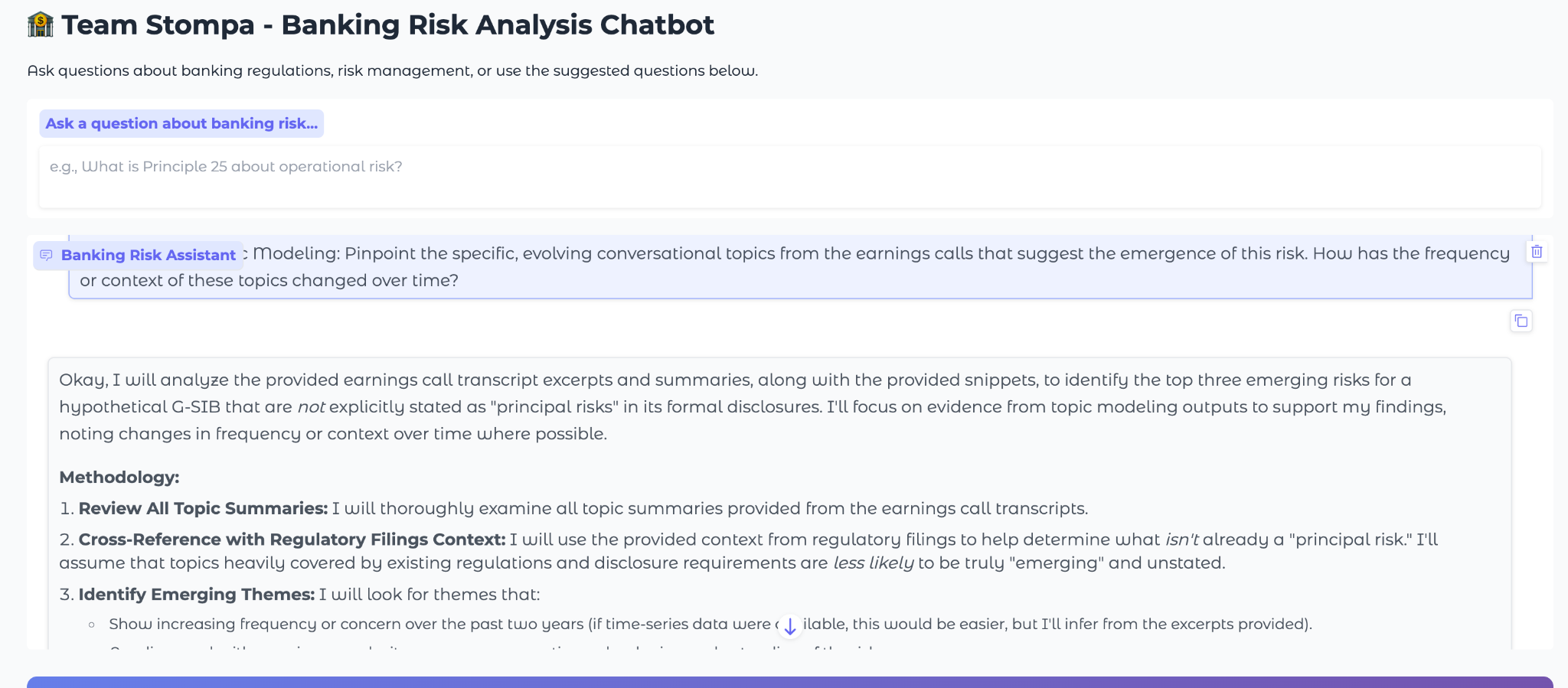
* Earnings call transcripts
* Basel Framework Docs.

The previous steps of the pipeline will be used to inform prompts based on identified topics and other indicators associated with risk. The retriever performs a semantic search to the vector database to find additional context that is relevant to the query. The prompt along with the retrieved data is then fed into the LLM to produce a more contextualised response.



Initially, for some queries the RAG was failing to accurately find context and reference accurate sources. A chunking strategy was then developed to increase the effectiveness of the semantic search to the vector database. Metadata has been added to each chunk to ensure they all contain key information about the source document. This has increased the ability of the pipeline to provide contextual responses for some prompts.

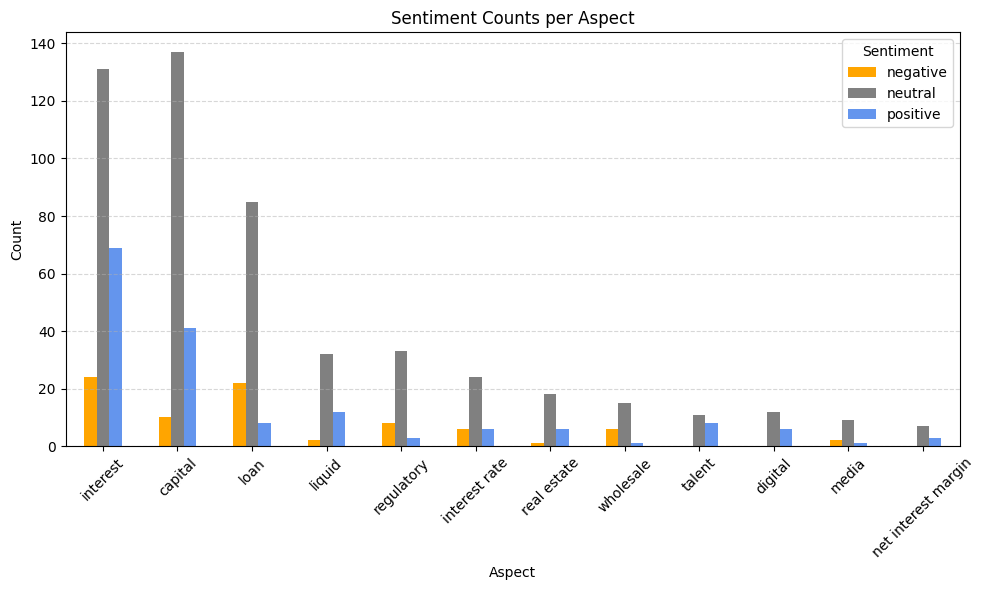
The final RAG has been integrated into a virtual assistant chatbot that can be queried and provide contextual responses including references to the relevant regulatory framework. The virtual assistant can also suggest questions that could provide relevant insight on financial risk. We also found that different LLMs performed differently on some tasks, as a result the chatbot contains functionality to query three different LLMs, which may provide different insights.



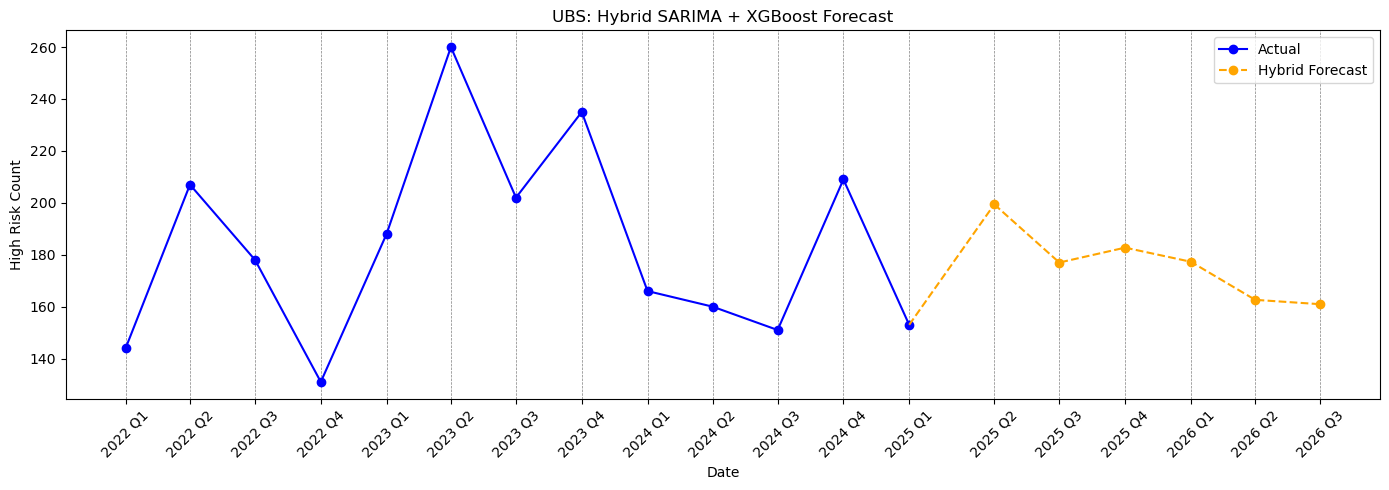
****

**3 Results**

Topic modelling revealed consistent areas of concern across G-SIBs. Net interest margin pressure, exposure to commercial real estate, capital adequacy, and digital transformation appeared as dominant themes. UBS and JPMorgan emphasised loan book quality and capital resilience, while Citibank focused on operational efficiency and liquidity management. Across banks, commercial real estate and cybersecurity risks also appeared frequently, often in connection with digital infrastructure investments and regulatory scrutiny.

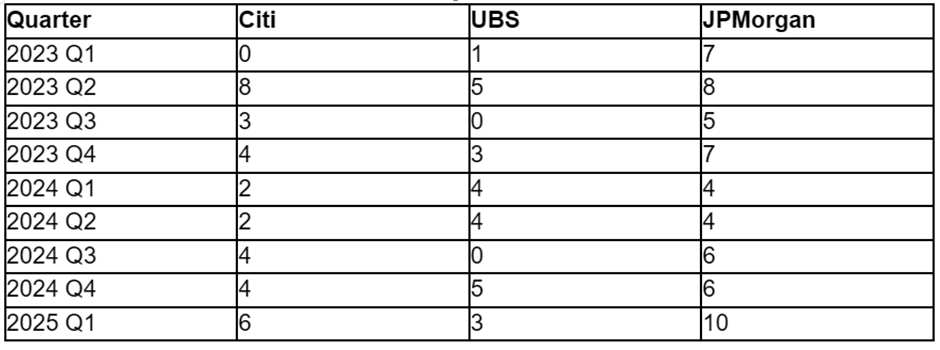
Aspect based sentiment analysis found the most frequently discussed aspects to be Loan Book Quality,Interest Rates, andCapital Adequacy. Overall sentiment was mostly neutral, though Loan Book Quality stood out with more negative than positive sentiment. Interest Rates showed a gradual rise in negativity over the past two years, while sentiment around Capital Adequacy has brightened slightly. In the last three quarters, concerns about Loan Book Quality have lessened, although it remains more negative than positive. Graphs of quarterly counts for the top three aspects can be found in appendix A.

Risk-Based Sentiment analysis shows that UBS exhibits the highest proportion of high-risk sentences during Quarterly Earnings Calls and maintains elevated levels through 2026 despite a slight decline. JP Morgan shows strong volatility with a projected peak in mid-2025, followed by a gradual decline. Citigroup demonstrates a steady recovery from a mid-year dip, trending upward into next year. Overall, the hybrid forecast model highlights UBS and JPMorgan as key institutions to monitor due to their elevated or fluctuating high-risk outlooks.



Question avoidance detection identified several instances where questions appeared to be deflected. These occurred most often during periods of external pressure or internal transition. The avoidance frequency fluctuates over time with an uptick in Q2 of 2023 and generally higher avoidance at JPMorgan.

****

****

Avoided questions primarily focused on:

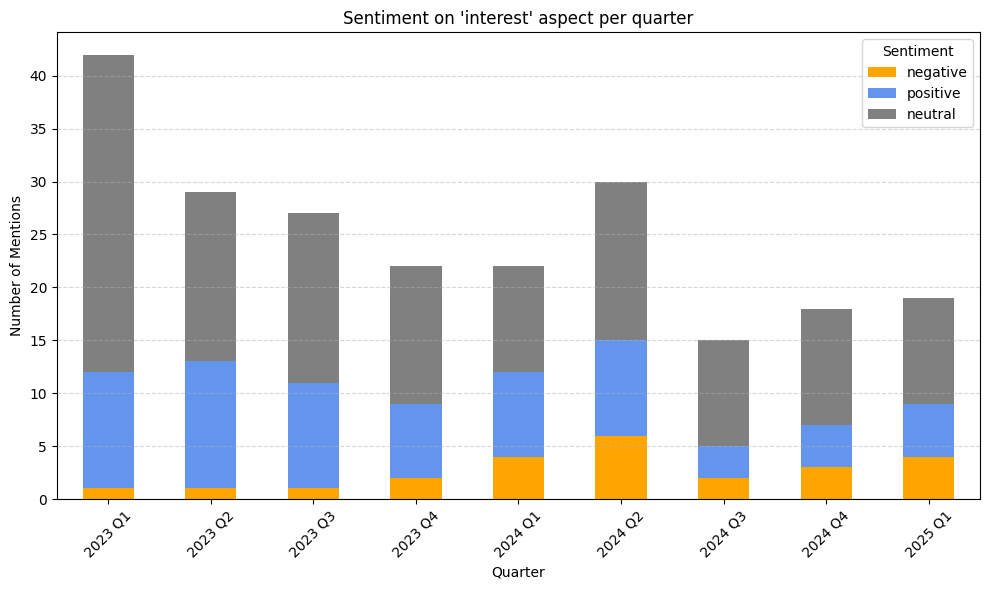
* **Citibank**: regulatory compliance and cost management
* **JPMorgan**: capital planning and macroeconomic resilience
* **UBS**: post-merger integration and capital return constraints

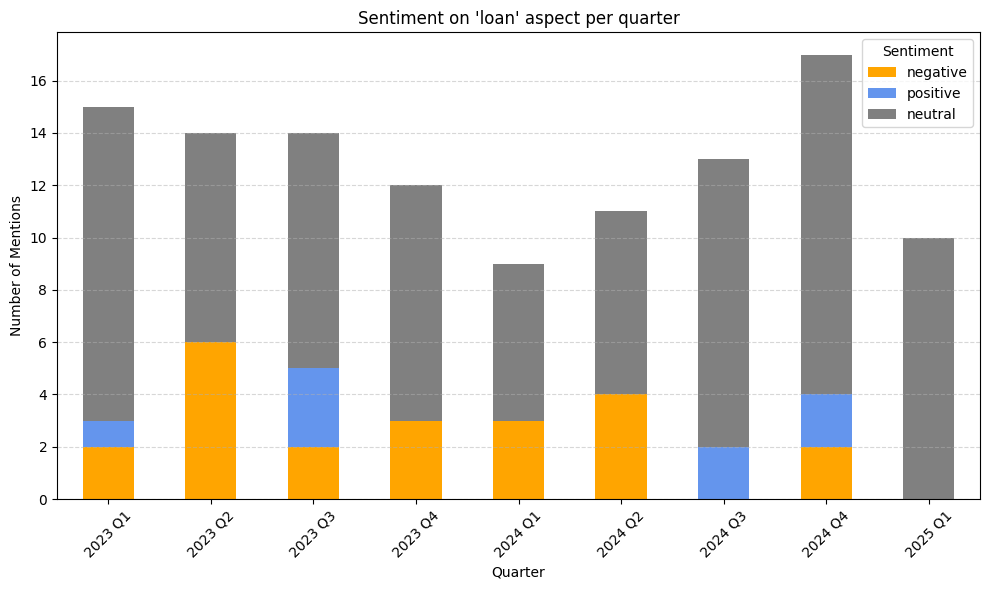
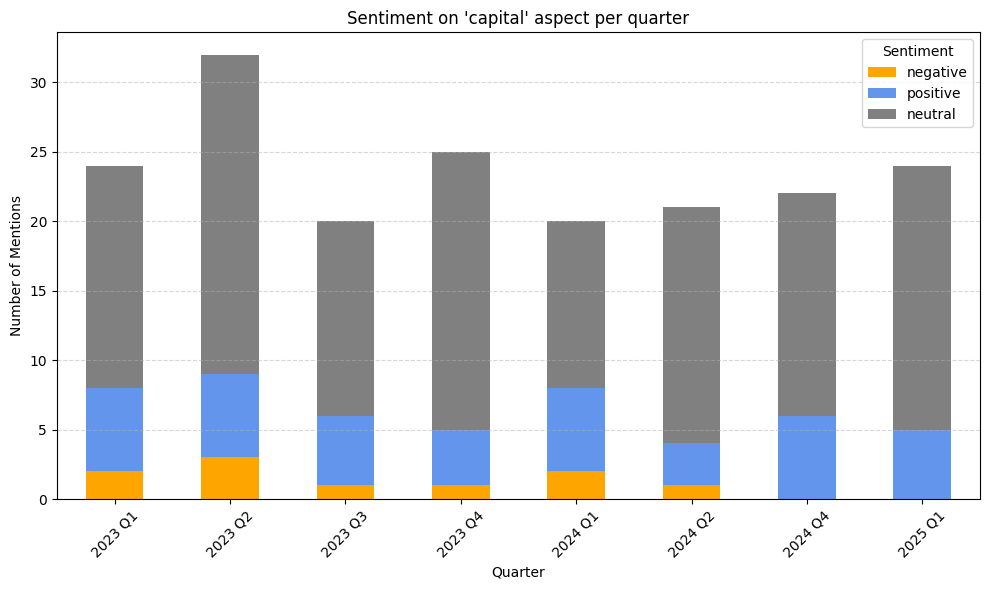
While these patterns stress a few common points, no single theme was consistently avoided across all banks.

The RAG chatbot provides a helpful tool for providing context to the user, meaning that specific outputs from the prior steps of the process could be analysed by someone without a specific finance background.

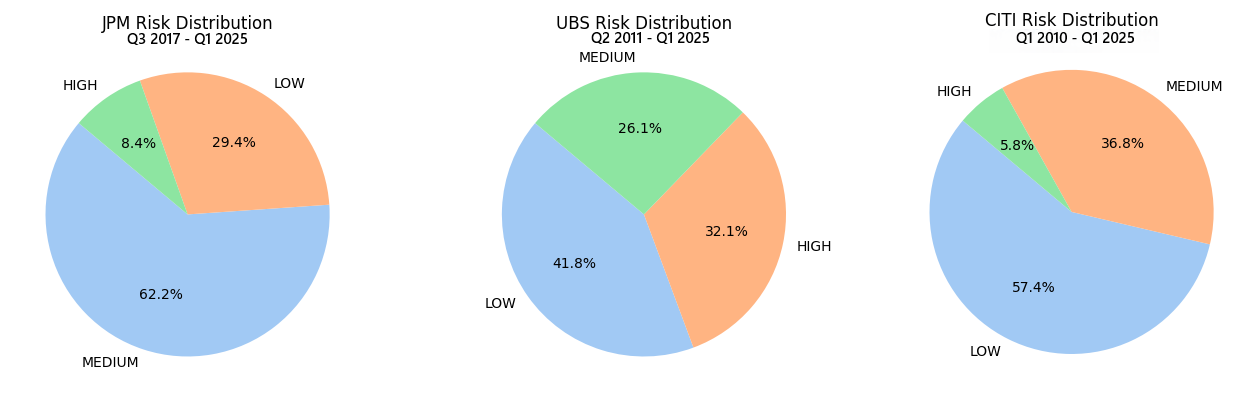
**Appendices**

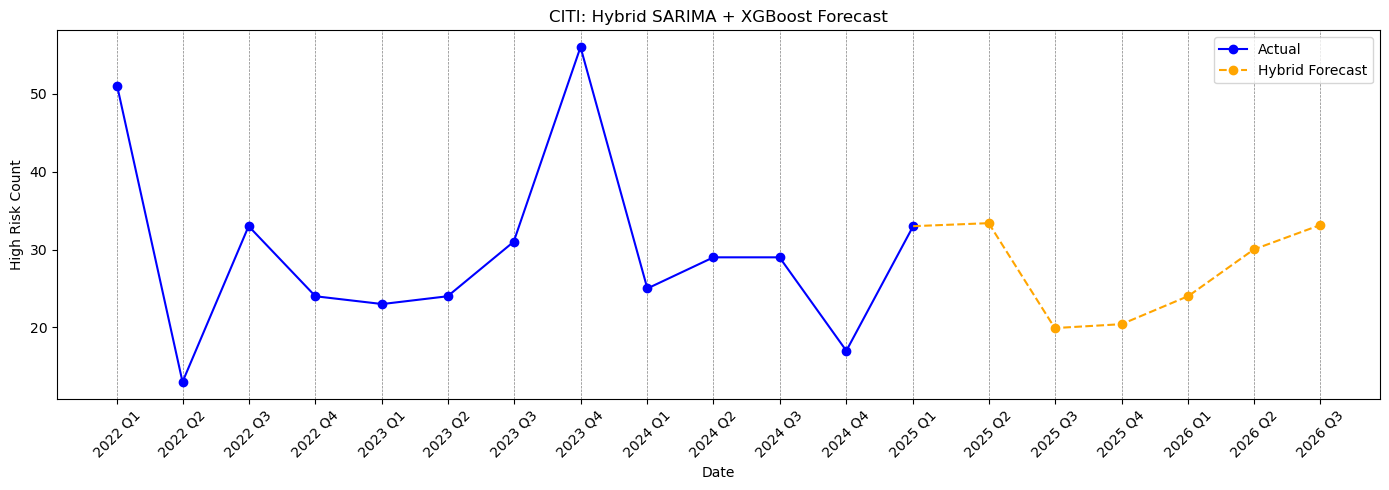
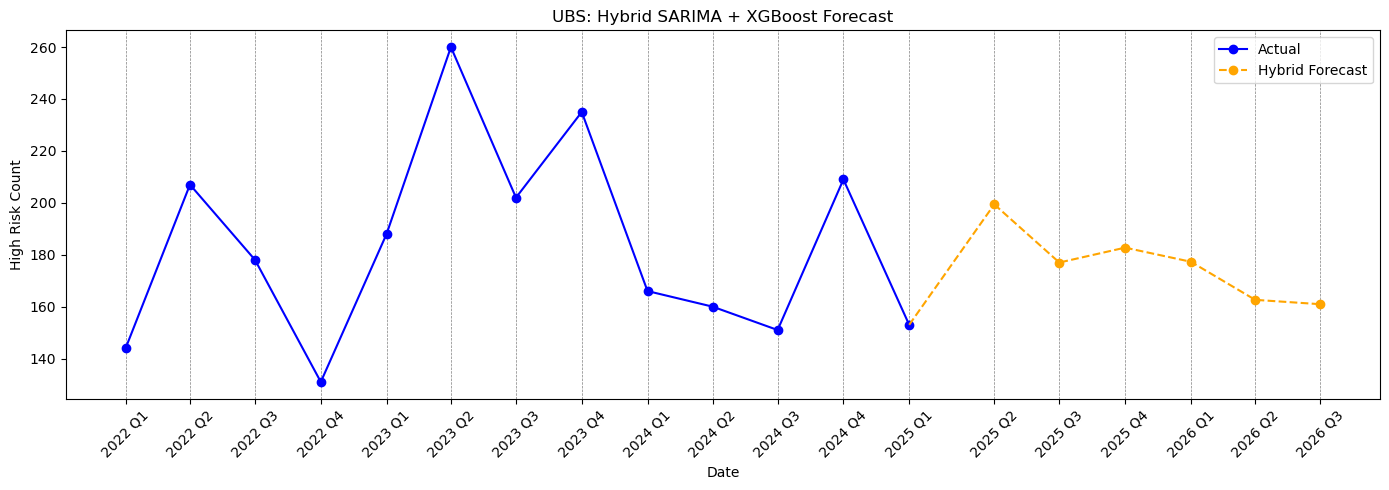
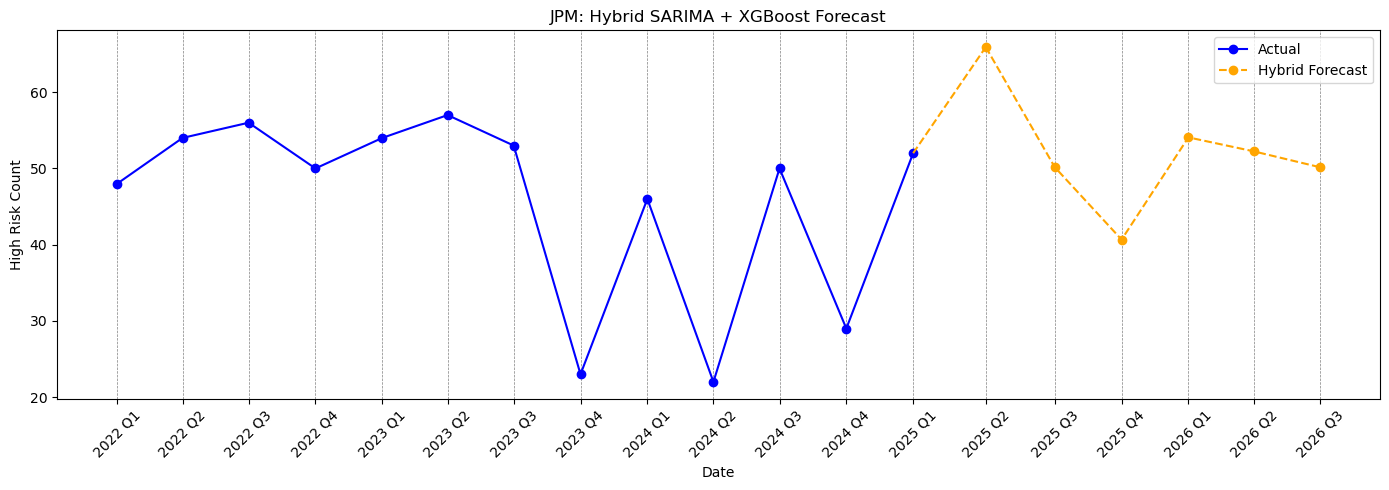
Appendix A





Appendix B





Appendix C

RAG Chunking Code - Block 1



RAG Chunking Code - Block 2



RAG Chunking Code - Block 3

