



Data Sentients: BoE Presentation



Indra, Anitta, Zarif, Rean, Nara



What if Al could help prevent the next financial crisis?



Contents

- 1. Goal & Strategy
- 2. Summary of Approach
- 3. MVP
- 4. Shortcomings & Future work



Section 1: Problem & Strategy



The business problem

- 1. **Challenge:** Bank of England PRA faces a significant challenge overseeing 1,500 financial institutions effectively
- 2. **Gap:** While numerical indicators are well established for risk monitoring, there is a crucial gap in leveraging qualitative data
- 3. **Opportunity:** Quarterly earnings calls with CEOs and Analysts are rich sources of untapped data that could reveal risk before numeric metrics
- 4. **Opportunity Gap:** Manual reviews are time-consuming and inefficient

We believe that modern NLP technology can help the Bank of England scale qualitative analysis and detect emerging risks before they appear in traditional financial metrics.



Our Goal

To improve the Bank of England's risk assessments of firms by applying advanced language models to analyse quarterly results transcripts.

If we analyse quarterly earnings call transcripts then we can identify signs of distress in a bank including early warning indicators that a bank is likely to collapse, uncovering insights that help understand a firm's stability and risk profile





Hypothesis

If we analyse quarterly earnings call transcripts then we can identify signs of distress in a bank including early warning indicators that a bank is likely to collapse, uncovering insights that help understand a firm's stability and risk profile



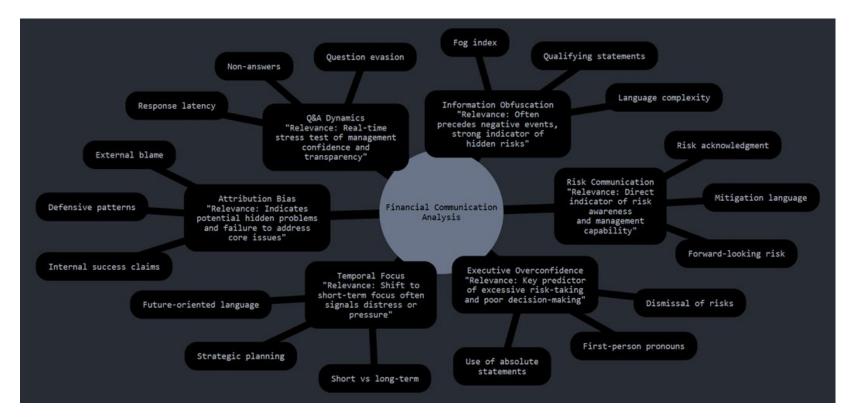


Strategy Overview - The Early Warning System





Beyond Financial Metrics: A Data-Driven Approach



Science of Executive Behaviour



- 1. **Risk Communication**: "The significance of existing risk disclosures, rather than new terms, increases before crises when validated against market behavior. Three conditions required." (Hanley & Hoberg, 2019)
- 2. **CEO Overconfidence**: "Banks with overconfident CEOs had higher leverage ratios and experienced greater losses during the 2007-2008 crisis." (Liu et al., 2020)
- 3. **Information Obfuscation**: "Companies with poor performance have difficult-to-read annual reports (higher Fog index) and show a 10% increase in complexity before negative events." (Li, 2008)
- 4. **Q&A Dynamics**: "Firms with highly scripted Q&As underperform in stock returns by 4-11% annually." (Lee, 2015)
- 5. **Attribution Patterns:** "57% of good news self-attributed by CEO, 100% of bad news externally attributed. This pattern preceded financial distress." (Brennan & Conroy, 2013)

ALERT Framework

Early Warning System





Attribution patterns

External blame | Internal credit | Learning resistance



Language complexity

Increased fog index | Reduced readability | Technical jargon



Executive over confidence

Risk underestimation | Excessive optimism | Reduced concern



Risk communication

Disclosure changes | Warning signals | Pattern shifts



Temporal changes in Q&A

Spontaneity loss | Response patterns | Script adherence

Bank Failure Risk

(Da Costa, 2024). Unpublished.



Literature Review

Beyond Financial Indicators



Behavioral Framework

5 Key Themes

Communication Markers

Linguistic Pattern Mapping

External Blame, Internal Credit, Speech Complexity, etc.



NLP Dictionary

Implementation Vocabulary

Specific words, phrases, patterns for analysis

```
external blame indicators = {
    'market conditions': [
        "market conditions",
        "economic environment",
        "industry headwinds",
        "external factors",
        "macro challenges"
    'external_forces': [
        "regulatory changes",
        "government policies",
        "competitive pressures",
        "outside forces"
    'passive voice': [
        "was affected by",
        "were impacted by",
        "resulted from",
        "due to circumstances"
```

```
internal_credit_indicators =
    'leadership claims': [
        "I decided",
        "I implemented",
        "my strategy",
        "my direction",
        "I initiated"
    'personal achievement': [
        "my track record",
        "my experience",
        "my leadership",
        "under my guidance"
```

```
learning resistance indicators =
    'dismissal': [
        "as we've always done",
        "our proven approach",
        "traditional method",
        "stick to our strategy"
    'repetition': [
        "once again",
        "as before",
        "maintain our course",
        "continue same approach"
```



Assumptions



- CEO language declines before financial metrics
- Quarterly Earnings Calls contain 'distress' signals
- NLP can detect these 'distress' signals
- Quarterly calls have sufficient frequency for detection of change over time



- Credit Suisse provides a good baseline distress case study
- Business model similarity
- Three banks provide good validity of data



Section 2: Summary of Approach



Approach

- Considered banks based on their risk profiles and operational similarities.
- 3 broad categories for the banks :
 - Distressed
 - Stable
 - Volatile
- Credit Suisse from distressed category was our initial MVP focus.
- Insights gained from the MVP influenced the bank selection strategy for the next stages:
 - Credit Suisse (Distressed) + SVB (Distressed) + Nomura (Volatile)
- 2 additional banks to be explored if there is time (Stable, Positive)





Bank	SVB (USA)	Credit Suisse (Europe)	Nomura (Asia)
Summary	Concentrated tech-sector risk, liquidity crisis, and panic triggered global instability.	Reputation issues, scandals, and liquidity problems amplified crisis impact.	Geographically diversified, strong risk management, conservative lending, and benefited from low-rate stability.

- > Trend: The 3 banks allow us to capture the industry trend rather than the region specific trend.
- Interrelation: Credit Suisse and Nomura both suffered heavy losses due to Archegos collapse highlighting Interrelation of firms with one recovering, while the other did not.
- Crisis Responses: SVB's liquidity issues and poor diversification are similar to Credit Suisse's problems, while Nomura responded better after Archegos with improved oversight and risk controls.
- A volatile bank, like Nomura, that performed poorly could have been on the brink of collapse. So selecting these 3 banks we are trying to find out were there any early warning signs before the losses? On the flip side, are there early indicators showing a higher risk appetite and better performance?



Data Gathering

 We have gathered relevant earning call transcripts of Credit Suisse, SVB and Nomura from "Seeking Alpha". The transcripts are stored in our repository in raw format for further analysis.





Data Preprocessing

Implemented the ELT pipeline to efficiently extract, load, and transform earnings call transcripts in PDFs, enabling seamless analysis and model integration.

Extract: Used PDFplumber to retrieve earnings call transcripts, storing the text in an intermediate format while preserving document structure.

Load: Stored the raw data in a structured repository, organizing it into columns (e.g., Speaker, QA dialogues, Quarter) for batch or real-time access as required.

Transform: Cleaned and refined the text data to normalize (e.g., lowercase, removed punctuation), segmented by speaker/quarter/QA dialogues/presentations, summarised transcripts using NLTK and prepared it as a structured dataset ready for deeper analysis.



LM Evaluation

Selection Criteria & Metrics

- Key performance metrics: Reasoning, contex window/token limit, specialisation, resource efficiency
 - o General purpose SLM with RAG vs General/Financial purpose LLMs with RAG
- 2 Instruct LMs to compare for MVP:
 - Model 1: microsoft/Phi-3-mini-4k
 - Model 2: <u>Phi-3.5-MOE-128k</u>

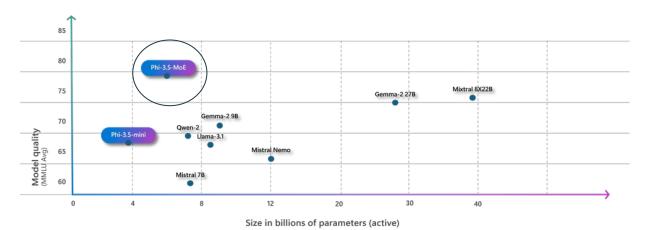






- Model 1: Phi-3-mini-4k Can be used with the RAM limitations on the free version of google colab
- Model 2 MVP: microsoft/Phi-3.5-MOE-instruct performs exceptionally well compared to larger models for its size

Phi-3.5 Quality vs Size in SLM



Model Selection & Performance Comparison



- Model for next gen work: Gemini-1.5-Flash-latest Ranked #4 on RepoQA benchmarking
- Scalable, multimodal support, requires fine tuning for financial domain

#	Model	Claimed Ctx	.py (%)	.cpp (%)	.rs (%)	.java (%)	.ts (%)	∑ Avg. (%)
1	<u>₩ gpt-4o-2024-05-13</u>	128k	95	80	85	96	<u> </u>	90.6
	<pre> gemini-1.5-pro-latest </pre>	1000k	91	81	91	94	△ 96	90.6
	<u>₩ claude-3-opus-20240229</u>	200k	93	83	88	<u> </u>	94	90.6
4	gemini-1.5-flash-latest	1000k	93	79	87	94	<u></u> 497	90.0
5	claude-3-sonnet-20240229	200k	88	81	85	△ 92	91	87.4



Model Selection & Performance Comparison

- Gemini flash performs the best compared to the models shown
- Does not require RAG or langchain with its large context length depending on number of banks and quarters

Category	Benchmark	Phi-3.5- mini- instruct	Mistral-7B- Instruct- v0.3	Mistral - Nemo-12B- Ins-2407	Llama- 3.1-8B- Ins	Gemma- 2-9B-Ins	Gemini 1.5 Flash	GPT-40- mini-2024- 07-18 (Chat)
	Arena Hard	37	18.1	39.4	25.7	42	55.2	75
Popular aggregated	BigBench Hard CoT (0-shot)	69	33.4	60.2	63.4	63.5	66.7	80.4
benchmark	MMLU (5-shot)	69	60.3	67.2	68.1	71.3	78.7	77.2
,	MMLU-Pro (0-shot, CoT)	47.4	18	40.7	44	50.1	57.2	62.8
	ARC Challenge (10-shot)	84.6	77.9	84.8	83.1	89.8	92.8	93.5
	BoolQ (2-shot)	78	80.5	82.5	82.8	85.7	85.8	88.7
	GPQA (0-shot, CoT)	30.4	15.6	28.6	26.3	29.2	37.5	41.1
	HellaSwag (5-shot)	69.4	71.6	76.7	73.5	80.9	67.5	87.1
Reasoning	OpenBookQA (10-shot)	79.2	78	84.4	84.8	89.6	89	90
	PIQA (5-shot)	81	73.4	83.5	81.2	83.7	87.5	88.7
	Social IQA (5-shot)	74.7	73	75.3	71.8	74.7	77.8	82.9
	TruthfulQA (MC2) (10-shot)	64	64.7	68.1	69.2	76.6	76.6	78.2
	WinoGrande (5-shot)	68.5	58.1	70.4	64.7	74	74.7	76.9
Multi-	Multi-lingual MMLU (5-shot)	55.4	47.4	58.9	56.2	63.8	77.2	72.9
lingual	MGSM (0-shot CoT)	47.9	31.8	63.3	56.7	76.4	75.8	81.7
Math	GSM8K (8-shot, CoT)	86.2	54.4	84.2	82.4	84.9	82.4	91.3
	MATH (0-shot, CoT)	48.5	19	31.2	47.6	50.9	38	70.2
Long	Qasper	41.9	31.4	30.7	37.2	13.9	43.5	39.8
context	SQuALITY	24.3	25.9	25.8	26.2	0	23.5	23.8
Code	HumanEval (0-shot)	62.8	35.4	63.4	66.5	61	74.4	86.6
Generation	MBPP (3-shot)	69.6	50.4	68.1	69.4	69.3	77.5	84.1
Α	verage	61.4	48.5	61.3	61.0	63.3	68.5	74.9

Phi-3.5-mini Model Quality

Modelling approach - Sentiment Analysis



We have chosen **yiyanghkust/finbert-tone** as it is specifically fine-tuned for sentiment analysis in the financial domain and it was trained on earnings call transcripts and financial news.

Model	Token Limit	Best For	Key Strengths	Key Limitations
yiyanghkust/finbert-tone	512 tokens	Financial sentiment analysis	High accuracy for financial sentiment	Limited to sentiment, less flexible for other NLP
ProsusAl/finbert	512 tokens	Sentiment, Entity Recognition, and general Classification	Versatile in financial text understanding	Less refined sentiment precision compared to finbert-tone
GPT-2	1,024 tokens	Summarization, Text Generation, Conversational AI	Adaptable to many NLP tasks with fine-tuning	Not finance-specific, requires fine-tuning for accuracy



Modelling approach - Topic Modelling

The following two models were considered:

- BERTopic Generalised tool to discover underlying themes or topics from short, sparse, or diverse text.
- LDA model from Gensim Preferable for more structured, long documents. Easily interpretable and computationally efficient.

From the findings as stated above, Gensim seems more suitable for topic modelling earnings calls



Regulatory & Ethical Challenges





Equality Act 2010



Bank of England

- 1. **GDPR:** Any analysis of earnings call transcripts must ensure data minimisation, avoid personal profiling, and maintain strict data protection protocols even for publicly available information.
- 2. FCA: Analysis must be fair, transparent, and support market integrity while avoiding any potential market manipulation
- 3. **Equality Act 2010:** Analysis methodologies must be demonstrably free from demographic bias and avoid any discrimination across different speaker groups.
- **4. BoE Policy:** Models must be interpretable, documented, and align with existing risk frameworks while maintaining clear audit trails and validation processes

Limitations: This short-term project focuses on establishing foundational NLP analysis capabilities. Full regulatory compliance with experts would require additional development, comprehensive testing, and extensive validation that extends beyond the current project scope.



Proposed Pipeline Integration & Challenges



INTERGRATION POTENTIAL

BoE's risk monitoring frameworks could be enhanced through:

- Establishment of a of supervisory review process to sync with earning calls
- 2. Integration with existing BoE risk dashboards
- 3. Develop automated alert system for risk indicators

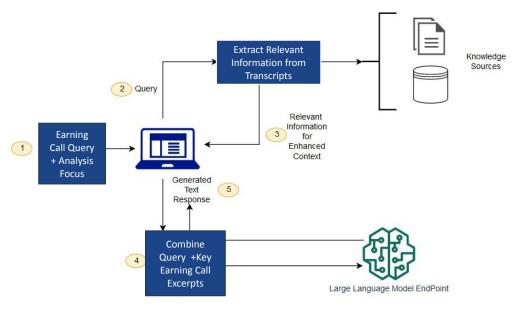
IMPLEMENTATION CHALLENGES

- 1. **Technical**: Integration with legacy systems & scalability
- 2. **Operational**: Process documentation, Staff training, Change management, resources for continuous improvement/monitoring
- 3. **Regulatory**: Model risk & data governance requirements, Validation frameworks, PRA Policy alignment



Use of Retrieval Augmented Generation (RAG)

Implementation of RAG (Retrieval-Augmented Generation) approach demonstrated exceptional results by effectively leveraging unstructured data, such as earnings call transcripts, to deliver precise and contextually relevant insights. These insights are invaluable for supporting informed regulatory decisions and enhancing strategic planning.





MVP

Information Summarisation, Sentiment Analysis, Topic Modelling



Responses from the AI Assistant Implemented using SLM and RAG

Small Language Model: microsoft/Phi-3-mini-4k-instruct

Welcome to the Earning Call Transcript-Based Risk Analyzer
An Early Warning System for Investors

Step 1: Select a Language Model

Available Language Models:
1. microsoft/Phi-3-mini-4k-instruct
2. microsoft/Phi-3.5-MoE-instruct
3. gemini-1.5-flash-8b
4. openai/gpt-4

Select a model (1-4) or press Enter for default [openai/gpt-4]:

Ouerv	Selection	Menu

- Analyze a single quarter
- 2. Compare two quarters of the same year
- 3. Compare two quarters of different years
- 4. Year-over-year comparison for all quarters
- 5. Analyze all quarters of a single year
- 6. Analyze sentiment for a financial year
- 7. Analyze combined presentation and dialog sentiment for a quarter
- 8. Compare sentiment trends between presentation and dialog
- 9. Generate aggregated quarter summaries with sentiment
- 10. Exit

Select an option (1-6):



Responses from the AI Assistant Implemented using SLM and RAG

Small Language Model: microsoft/Phi-3-mini-4k-instruct

_____ Ouery Selection Menu 1. Type in any question 2. Compare two quarters of the same year 3. Compare two quarters of different years 4. Year-over-year comparison for all quarters 5. Analyze all quarters of a single year 6. Analyze sentiment for a financial year 7. Analyze combined presentation and dialog sentiment for a quarter 8. Compare sentiment trends between presentation and dialog 9. Generate aggregated quarter summaries with sentiment Select an option (1-6): 1 2024-11-21 13:10:34.485 - INFO - User selected option: 1 INFO:PDFProcessor:User selected option: 1 Enter your query: Whats the overall sentiment in year 2020? Enter the bank name: Credit Suisse 2024-11-21 13:10:58,673 - INFO - Retrieving documents for filters: {'bank': 'credit suisse'}



[RESULT] Response from the AI Assistant:
The overall sentiment in 2020 for Credit Suisse, as described in the provided text, appears to be cautiously optimistic despite facing significant challenges. The bank reported a pretax income of CHF 3.5 billion and net income attributable to shareholders of CHF 2.7 billion for the full year. This indicates a positive financial performance. However, the bank also faced considerable headwinds, including a CHF 1.1 billion provision for credit losses and significant foreign exchange (FX) headwinds due to a weaker U.S. dollar. Despite these challenges, Credit Suisse achieved a 6% increase in adjusted pretax income.



Responses from the AI Assistant Implemented using LLM and RAG

Large Language Model: microsoft/Phi-3-MoE-4k-instruct

User: What is the overall sentiment for Credit Suisse in the year 2020?

Answer: The overall sentiment for Credit Suisse in the year 2020 appears to be Neutral (0). Despite facing challenges such as the need to record a pretax loss in the fourth quarter due to historic issues and litigation provisions, the bank also reported a pretax income of CHF 3.5 billion and a net income attributable to shareholders of CHF 2.7 billion for the full year. The bank also made progress in addressing legacy issues, strengthening risk and compliance teams, and investing in its businesses. Furthermore, it launched sustainability research and investment solutions, SRI, and entered 2021 with strong momentum, as evidenced by its best January in a decade. However, the bank also faced significant challenges, including a pretax loss in the fourth quarter and a difficult 2021 due to issues related to supply chain finance funds and Archegos matters. Therefore, while there were positive aspects, the challenges faced balance them out, resulting in a neutral sentiment.



Responses from the Al Assistant Implemented using LLM and RAG

Large Language Model: microsoft/Phi-3-MoE-4k-instruct

User: Summarise the stability of Credit Suisse for Q2 and year 2022?

Answer: Credit Suisse exhibited stability in Q2 and throughout the year 2022, despite facing challenges. The bank maintained a robust balance sheet, with a CET1 capital ratio of 14.1% and a Tier 1 leverage ratio of 7.7% at the end of Q4 2022. The bank also released more than \$3 billion in allocated capital from the Investment Bank by the end of the year, exceeding its target. Credit Suisse continued to execute its transformation strategy, which includes simplifying its structure, strengthening its core businesses, and investing for growth. The bank also made progress in restructuring its Investment Bank and creating CS First Boston. Additionally, Credit Suisse focused on improving risk management, compliance, and addressing legacy issues. The bank's capital and liquidity ratios remained strong, and it continued to pay dividends, demonstrating its financial stability. However, the bank faced challenges such as a net loss in Q4 2022 and a challenging macroeconomic and operating environment. Overall, Credit Suisse demonstrated resilience and a commitment to executing its strategic initiatives, which should position it for future growth.



Credit Suisse - Sentiment Analysis output

index	# Yea		A⊐ Q	uarte	er	A Bank		≜ Se	ction ·	·· 👌 Se	ntiment	# Pos	itivity i	# Neu	trality	# Neg	ativity	
()	2020	Q2			bank_crec	lit_suisse	Combi	ned	Neutra	al		38.18		58.18		3.64	
	1	2020	Q3			bank_crec	lit_suisse	Combi	ned	Neutra	al		30.0		70.0		0.0	
	# index	#	Year		∆ Qua	rter	A Bank		A Sec	tion ···	≜ Sentime	ent	# Positivit	y	# Neutra	lity	# Negativ	ity
		0	2	2020	Q2		bank_cred	it_suisse	Present	ation	Positive			60.0		40.0		0.0
		1	2	2020	Q3		bank_cred	it_suisse	Present	ation	Neutral			40.0		60.0		0.0
		# in	dex	#	Year	A □ Qua	rter {	Bank		∆ Section	on ···	A Sen	timent	# Po	sitivity	# Ne	utrality	# Negativity
					2020	Q2	i ja	oank_credi	t_suisse	Question-	and-Answer	Neutral			36.0		60.0	4.
					2020	Q3		oank_credi	t_suisse	Question-	and-Answer	Neutral			29.23		70.77	0.
			2		2020	Q4	1	oank_credi	t_suisse	Question-	and-Answer	Neutral			23.88		73.13	2.9
1			3		2021	Q1	i i	oank_credi	t_suisse	Question-	and-Answer	Neutral			20.34		77.97	1.6
			4		2021	Q2	9	o <mark>ank_</mark> credi	t_suisse	Question-	and-Answer	Neutral			12.5		82.14	5.3
			5		2021	Q3	1	oank_credi	t_suisse	Question-	and-Answer	Neutral			20.0		76.0	4.
			6		2021	Q4		oank_credi	t_suisse	Question-	and-Answer	Neutral			23.4		74.47	2.1
		1	7		2022	Q1		o <mark>ank_</mark> credi	t_suisse	Question-	and-Answer	Neutral			30.36		66.07	3.5
			8		2022	Q2	1	oank_credi	t_suisse	Question-	and-Answer	Neutral			25.0		75.0	0.
			9		2022	Q3		oank_credi	t_suisse	Question-	and-Answer	Neutral			23.08		73.08	3.8
			10		2022	Q4	1	oank_credi	t_suisse	Question-	and-Answer	Neutral			24.29		71.43	4.2

Transcripts → Summarised Transcripts

Result: Sentiments remain the same





NLTK - WordCloud

WordCloud - Quarterly - Original Transcript

WordCloud - Quarterly - Summarised Transcript

BERTopic

Original Transcript

BERTopic - Combined

BERTopic - Presentation

BERTopic - Question-and-Answer

Summarised Transcript

BERTopic - Combined

BERTopic - Presentation

BERTopic - Question-and-Answer

Gensim - LDA

Original Transcript

Gensim - LDA - Combined

Gensim - LDA - Combined - per quarter

Gensim - LDA - Presentation - per quarter

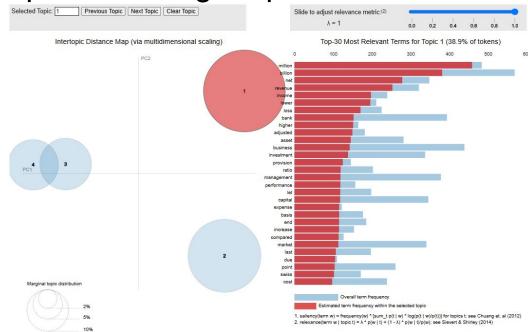
Gensim - LDA - Question-and-Answer - per quarter

Summarised Transcript

Gensim - LDA - Combined - per quarter

Gensim - LDA - Presentation - per quarter

Gensim - LDA - Question-and-Answer - per guarter



Year: 2022, Quarter: Q3, Coherence Score: 78.65%, Perplexity: -5.93

Topic #1: billion, capital, outflow, year, parent, period, investing, announcement, entity, region

Topic #2: funding, liquidity, announcement, reduction, u, today, october, spread, deposit, well

Topic #3: billion, well, around, region, announcement, strategy, next, funding, outflow, combination

Topic #4: seen, rumor, last, october, much, thanks, outflow, couple, certainly, affected







Validation

Concept Validation

- Client verification
- Market Analysis
- MVP

Technical Validation

- Historically backtested sentiment and topics on distressed banks across 3-12 month periods to identify early warning signals
- Behavioural markers to be created from peer reviewed research papers
- For we will validate RAG factual accuracy and insight accuracy
- In one key example, the team reviewed Q&A exchanges from Credit Suisse's 2022 Q4 earnings call, comparing our model sentiment scores with our own team ratings to verify report accuracy

0

Value Validation

- We will track false positives/negatives
- Do a peer comparison across banks
- o Final review with BoE will serve as review by financial experts / stakeholders



Improving Results - Topic Summarisation

To enhance topic summarization effectively, a structured approach incorporating continuous testing and strategic improvement is recommended.

This refined approach includes:

- Prompt Refinement: Regularly test and update prompts for clarity and relevance, adapting to nuances in financial language.
- LangChain Workflow Tuning: Optimize workflows by adjusting task sequencing and incorporating feedback for smoother information flow.
- RAG Framework Enhancement: Validate and update retrieval sources fine-tune generation for precise, fact-based outputs.

Improving Results



Topic Modelling

- Word cloud to visualise frequently used words
- Gensim LDA model: Hyperparameters (number of topics) are tuned using grid search to get the best topic coherence score and models are evaluated using Perplexity
- Summarised the transcripts to improve interpretability

Sentiment Analysis

- Per Comments from Company Participants
- Combination of Text Summarisation and chunking
- Per Quarter
 - Analysis of Presentation, Question & Answer and Combined sections
 - Analysis of Original Transcript vs Summarised Transcript



Shortcomings & Future work

- Incorporate sentiment analysis of multimodal data computer vision & voice data (including tone, pauses and body language)
- *Banks have subsidiaries operating in the UK subject to PRA regulation, hard
 to separate the subsidiary from the main holding company/entity
- Data Incorporate data from other qualitative sources, including annual returns and mandatory filings with regulators, for example.
- Explore how NLP can be combined with stock market signals
- Fully incorporate behaviour and linguistics research
- Additional Banks
- Time scales Timeframe and resources were only viable for developing MVP



Subsidiaries

Bank	Operates in the UK?	Subject to BoE PRA Regulation?	UK Subsidiary Name(s)
Silicon Valley Bank (SVB)	Yes	Yes	Silicon Valley Bank UK Limited: A UK subsidiary primarily serving the UK's tech and innovation sectors. Acquired by HSBC in March 2023.
Credit Suisse	Yes	Yes	Credit Suisse International: UK-based subsidiary focusing on investment banking and trading operations.
Nomura Holdings	Yes	Yes	Nomura International plc: London-based subsidiary handling European investment banking and capital markets operations.
JP Morgan Chase	Yes	Yes	J.P. Morgan Securities plc and J.P. Morgan Europe Limited: Key UK entities covering investment banking, asset management, and clearing.
HSBC	Yes	Yes	HSBC UK Bank plc: A ring-fenced bank for UK retail and commercial banking. Also operates HSBC Holdings plc, the group's global headquarters.



Thank you