

# Uncovering Risk in Bank Earnings Calls

Using NLP to detect **early warning indicators of risk** in Q&A transcripts

## Team FinSight

Lauren Brixey

Jerome Ahye

Anastasia Krivenkovskaya

Heidi Santos

Jyoti Rathod

Parush Parushev



UNIVERSITY OF  
CAMBRIDGE

# Project Background

## Problem

**Quarterly reports → structured but miss tone, sentiment and implicit concerns**

Earnings call Q&A exchanges often reveal hidden risks and evasiveness

PRA currently lack tools to systematically monitor transcripts at scale

## Business Objective

**Investigate whether NLP can detect early warning indicators of risk in Q&A transcripts**

Deliver a reproducible NLP pipeline to enhance supervisory monitoring capability

**Intended Audience:** Bank of England Data Science Team

# Pipeline

## Q&A Transcripts

J.P.Morgan  
HSBC ↔



## Preliminary Preprocessing

## Task Specific Preprocessing



## Topic Modelling



## Sentiment Analysis



## Evasion Detection

**Proof Of Concept:** J.P. Morgan  
**Benchmarking:** HSBC



## Summarisation

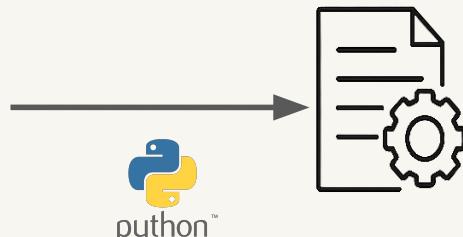


## PRA-ready Report

# Data & Preprocessing

Converting transcripts to standardised inputs for analysis

	Q
Christopher McGratty Analyst, Keefe, Bruyette & Woods, Inc.	
Great. Thanks for having me.	
	A
Jeremy Barnum Chief Financial Officer, JPMorgan Chase	
Just wanted to welcome you to the call before we get started.	
	Q
Christopher McGratty Analyst, Keefe, Bruyette & Woods, Inc.	
Great. Thanks so much. Relative to three months ago, there's a lot of optimism on financial deregulation and really going to break in the bank's favor. I'm interested if you agree, number one, with this optimism and anything specifically you could point to? And secondarily, on capital. I'm interested in what's on or off the table in terms of uses of capital? What do you need to see from the macro, regulatory? And how should we be thinking about the timing? Thank you.	
	A
Jeremy Barnum Chief Financial Officer, JPMorgan Chase	
Let me take the second part of that first. So, we have our centered capital hierarchy that we recite a lot, and I don't want to bore you by restating it again. But I think it's important because it does serve as a guide in this context, right? So, we deploy our capital against organic and inorganic growth, and we ensure a sustainable dividend. And with what's left, we do buybacks. And so, we've talked about how excess capital's earnings in store.	
You saw this quarter that we actually had some financial resource usage that came through and actually meant that, despite keeping the buybacks relatively constant and having the organic capital generation be relatively constant, the CET1 ratio dropped a little bit as a function of increased usage organically showing up in RWA expansion. So, we're doing what we want to do. But clearly, it is a big amount of excess and that does mean that everything is on the table as it always is, and that includes potentially inorganic things.	
Now, obviously, that needs to be done carefully. I think acquisitions have a high bar, both financially, strategically, and importantly, in some cases, culturally. And we also need to think carefully about things that work outside the regulated perimeter, might not work inside the regulated perimeter as well. We have learned some lessons. We don't want to overlearn those lessons.	



Field	Values
section	'Presentation', 'qa'
question_number	NULL, 1, 2, 3 ...
answer_number	NULL, 1, 2, 3 ...
speaker_name	'Jeremy Barnum'..
role	'Chief Financial Officer', ..
company	'JPMorganChase'
content	'<Verbatim Q or A>'
year	'2025'
quarter	'Q2'
is_pleasantry	TRUE/FALSE

# Approach

**Objective:** Topic modelling aligned with PRA regulatory themes.

Approach	Strengths	Weaknesses
LDA	Probabilistic, interpretable	Poor semantic capture, preprocessing-heavy
Top2Vec	Embedding-based, captures meaning	General-purpose embeddings, less control
BERTopic	High coherence, flexible, guided modelling	Needs tuning, embedding choice critical

## Model optimisation strategy:

- Domain-relevant embeddings → semantic comprehension
- Weighted PRA risk keywords → guided clustering

# What themes dominate analyst Q&As?

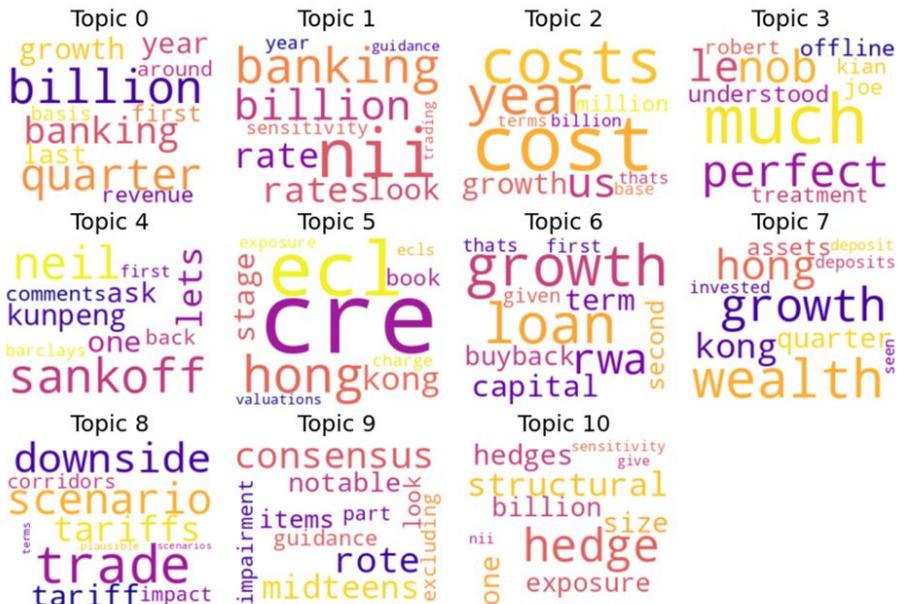
JP Morgan EC Topics	HSBC EC Topics
<ul style="list-style-type: none"><li>• Securities and market trends</li><li>• Deposit growth and interest rates</li><li>• Private lending and credit market</li><li>• Financial strategies and plans</li><li>• Macroeconomic factors and public sentiment</li><li>• Investor sentiment and market risk</li><li>• Customer interactions and relationships</li><li>• Macroeconomic sentiment (unemployment and recession)</li></ul>	<ul style="list-style-type: none"><li>• Geopolitical &amp; market risks (Hong Kong)</li><li>• Cost sensitivity and economic growth</li><li>• Operational cost management</li><li>• Strategic outlook and growth plans</li><li>• Inflation and macroeconomic factors</li><li>• Structural and hedge risk management</li></ul>

# How does one bank's thematic profile compare to peers?

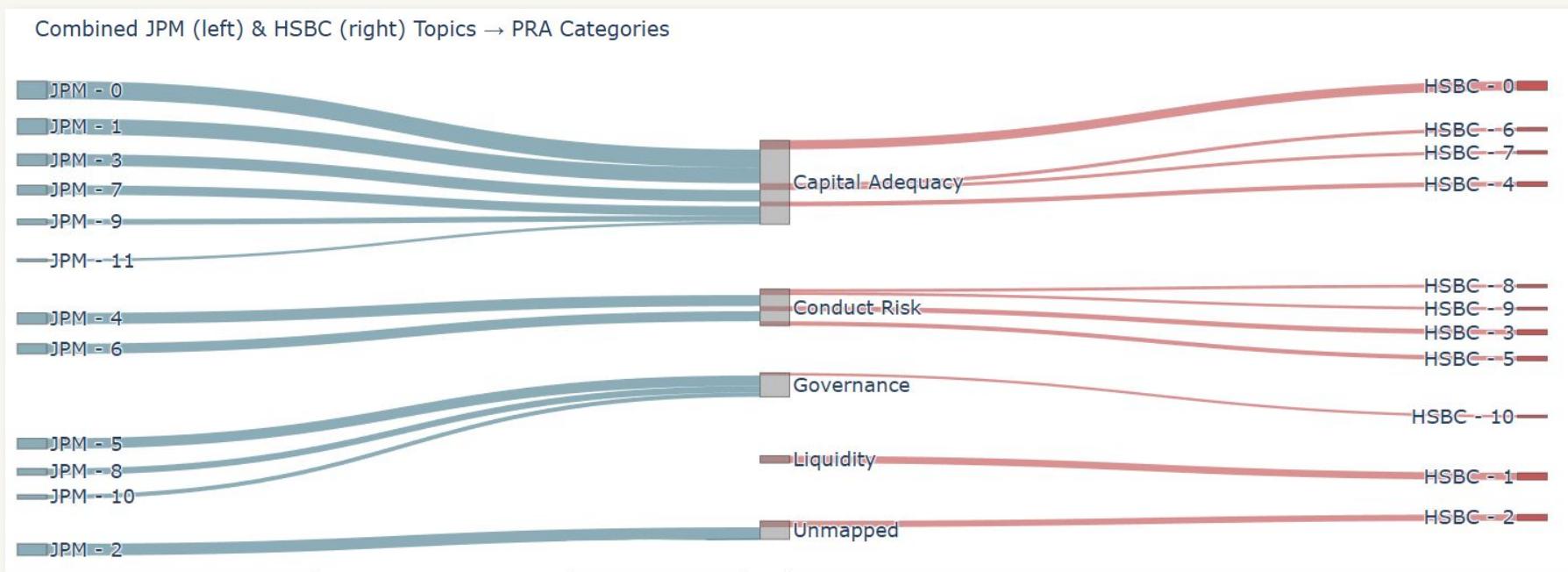
JP Morgan Word Clouds representation



HSBC Word Clouds representation

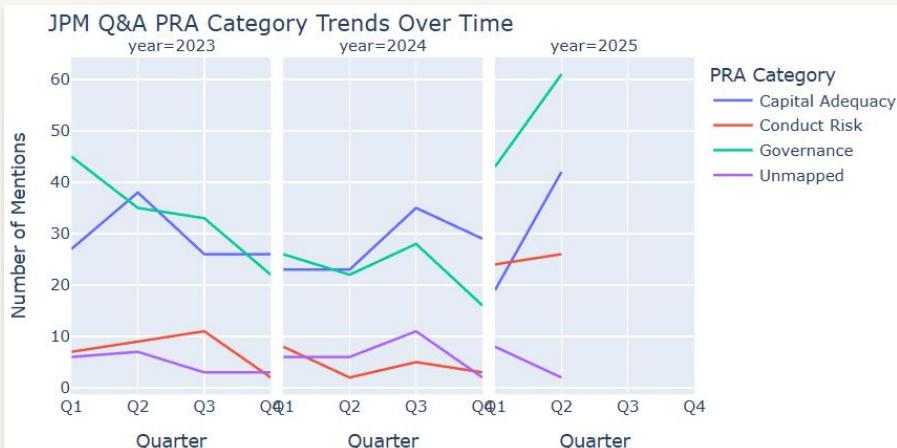


# How do these themes map to PRA risk categories?



# Do analysts revisit the same risks across quarters?

J.P.Morgan



HSBC



## Model Set-Up

- **4 transformer models** – FinBERT (2 variants), DistilRoBERTa, CardiffNLP RoBERTa
- **3-class sentiment analysis** (*positive / neutral / negative*)
- **Sentence-level segmentation** (2,785 Q&A samples; JPM: 1,505 | HSBC: 1,280)
- **Speaker role differentiation** (bankers vs analysts)

Model	Pre-trained Agreement (%)	Financial Alignment (%)	Fine-tuned Accuracy (%)	Composite Score
FinBERT (ProsusAI)	71.9	71.8	16.7	0.582
FinBERT (Yiyangkust)	70.8	69.7	20.8	0.577
CardiffNLP RoBERTa	68.1	60.9	66.7	0.526
DistilRoBERTa	61.0	71.1	62.5	0.472

- Chose **DistilRoBERTa** as the most stable model after cross-model evaluation
- Integrated **manual validation** (40 samples) and fine-tuning for class balance
- **Macro-F1** as main metric; stratified 80/20 train-test split
- **Fine Tuning** with learning rate 2e-5, 5 epochs, early stopping

# Do bankers and analysts show diverging sentiment?

Clear divergence detected between roles

Role	Negative (%)	Neutral (%)	Positive (%)	Sentiment
Analysts	4.5	86.8	8.6	cautious
CFOs	6.8	73.6	19.6	optimistic
Executives	6.8	84.3	8.9	balanced



Role-based sentiment divergence offering early signals of potential risk bias for PRA monitoring

# Has the tone shifted over time?

Highly stable sentiment patterns over time (Q1 2023 - Q2 2025)

Bank	Model	Average Tone Shift (%)
JPM	DistilRoBERTa	0.8
	CardiffNLP	4.7
HSBC	DistilRoBERTa	0.2
	FinBERT (ProsusAI)	6.5

**Stable macro environment  
(no crisis events for stress variation)**

# How does one bank's tone compare to peers?

Distinct communication styles detected

Bank	Model	Cross-bank Variation (%)	Top Topics
JPM	FinBERT (ProsusAI)	24.9	Market (8.4%) Revenue (3.6%) Regulatory (1.9%)
HSBC	FinBERT (Yiyanghkust)	13.4	Revenue (9.4%) Market (8.1%) Capital (2.3%)



Cross-bank divergence reflects regional and strategic communication cultures, valuable for peer benchmarking and supervisory comparison

# Are divergences systemic or firm specific?

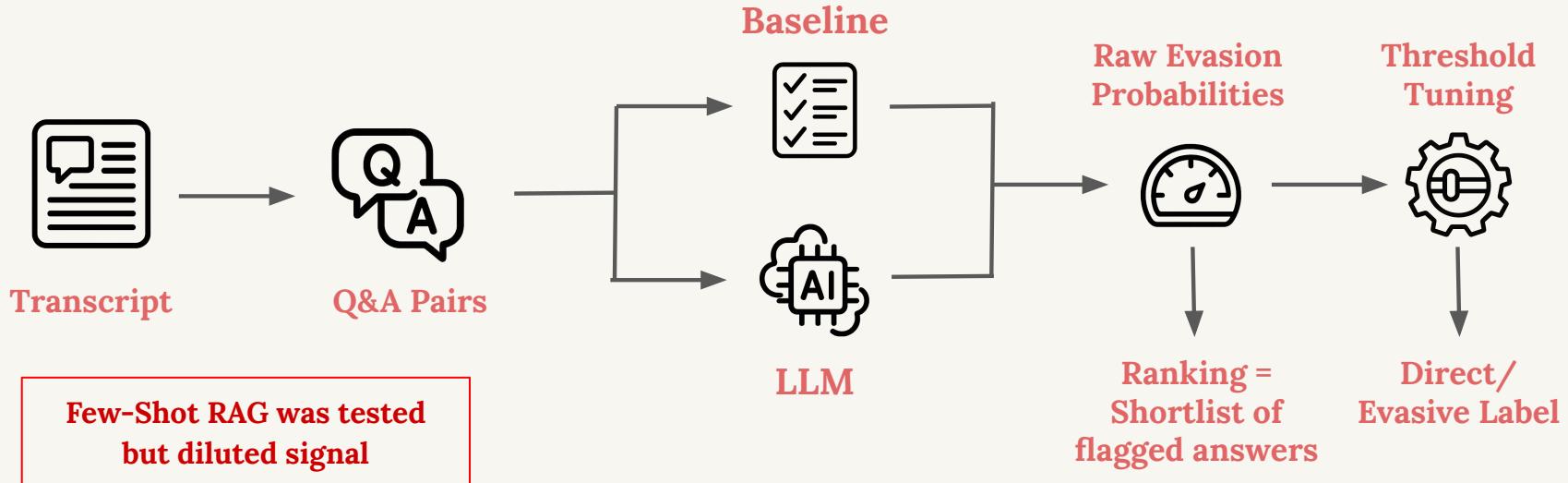
Firm-specific divergences confirmed

Evidence Supporting	Evidence Against	Limitations
<ul style="list-style-type: none"><li>• High consistency: 87.5–88.9 % cross-bank agreement</li><li>• Distinct style: 13.8 % variation = corporate culture (US vs UK)</li><li>• Independent tone movements: no synchronized Q1-Q2 shifts</li></ul>	<ul style="list-style-type: none"><li>• Stable tone across institutions over time</li><li>• Similar risk language intensity (~5–10 %) = normal range</li><li>• 68 % inter-model agreement = high reliability</li></ul>	<ul style="list-style-type: none"><li>• Only 2 banks (need 10–15 for systemic testing)</li><li>• Limited coverage (2 years, no crisis events)</li><li>• No correlation testing vs market indices or credit spreads</li></ul>

Expand dataset to >10 banks across 3–5 years to confirm systemic patterns

# Approach

Detect evasive answers so regulators know where to look



# Can we flag instances of banker evasiveness?

## Proof of Concept using simple probes

### Probe 1

**Q: What was CET1 at year-end?**

**A: CET1 ratio was 13.4%, up 40 bps YoY.**

### Probe 2

**Q: When will you release site losses?**

**A: We don't provide those numbers at this time.**



Models behaved as expected

Probe	Model	Probability of evasion (%)
1	DeBERTa	40.7
	Avg LLM	21.2
	Baseline	23.1
	Blended (base + DeBERTa)	34.1
2	DeBERTa	72.6
	Avg LLM	78.1
	Baseline	59.1
	Blended (base + DeBERTa)	67.6

# Can we flag instances of banker evasiveness?

Real Q&A transcripts are less clear-cut, making detection subjective and technically challenging

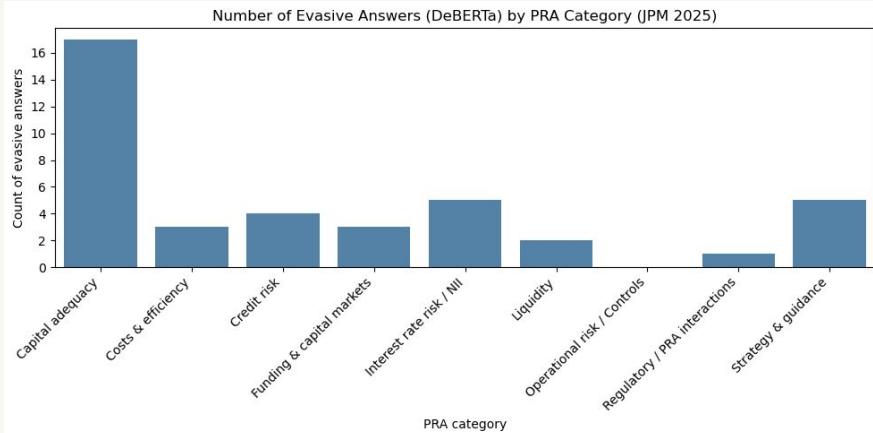
J.P.Morgan  
Test Set

Model	Evasive Detected (%)	Evasive Missed (%)	Direct Detected (%)	Direct Missed (%)	P@25% (%)
Baseline	93.8	6.2	12.6	87.4	10.7
DeBERTa	68.8	31.2	46.3	53.7	14.3
Avg LLM	62.5	37.5	38.9	61.1	17.9
Blended (base + DeBERTa)	75	25	19.7	82.1	17.9
RoBERTa					21.4

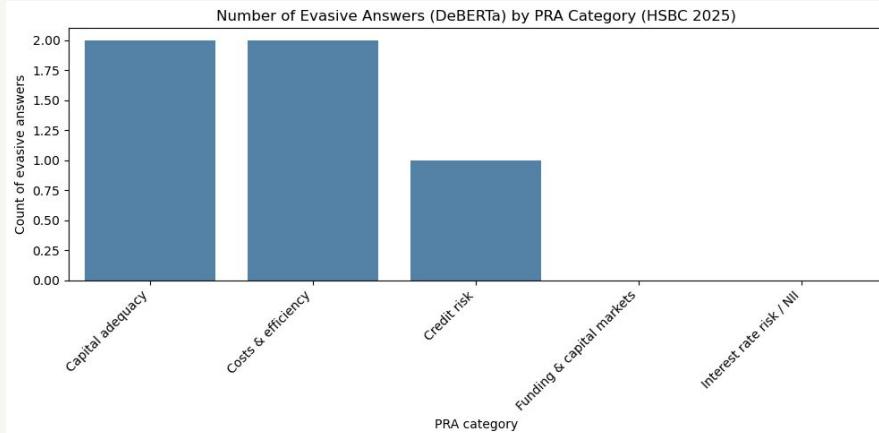


# Can we flag instances of banker evasiveness by PRA category?

J.P.Morgan



HSBC



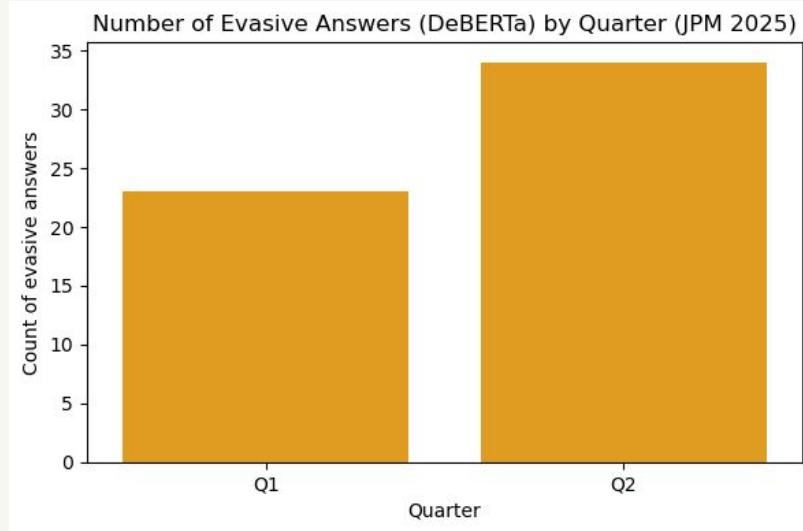
Capital adequacy



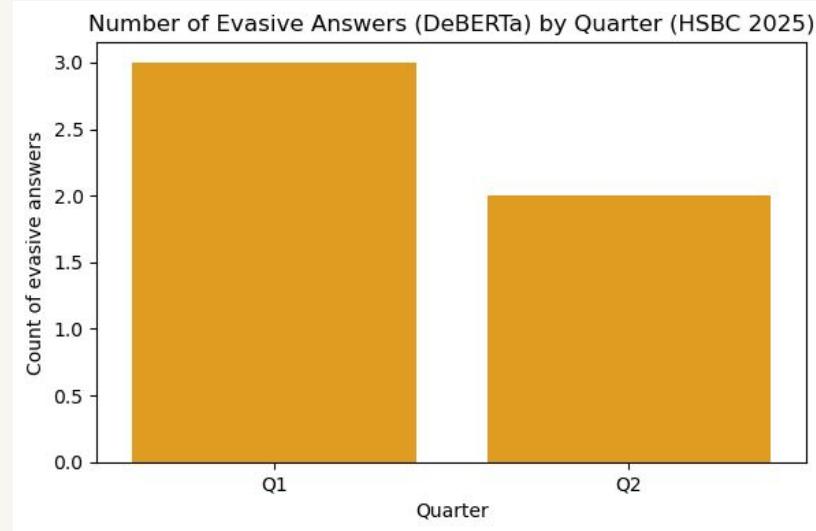
Capital adequacy &  
Costs and Efficiency

# Can we flag instances of banker evasiveness by Quarter?

J.P.Morgan



HSBC



# Approach

**Goal:** Can we synthesise concise regulator-ready briefings, at scale?

## Inputs

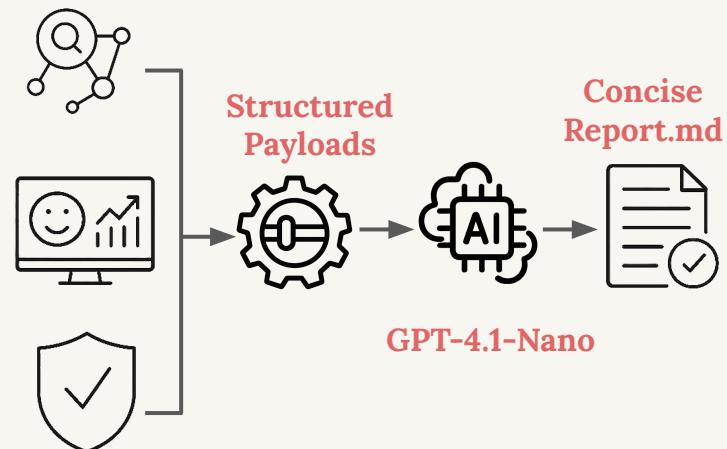
Topic, Sentiment, Evasion CSVs processed into LLM friendly ‘payloads’

## Model

GPT-4.1-nano synthesis (big central box with “Risk-Aligned Summaries”)

## Outputs

PRA-ready headlines + Q3 watch-outs



# Can we generate concise summarise aligned to risk categories?

## Themes:

Capital Adequacy is the dominant theme so far in 2025 for both JPMorgan and HSBC, with a slight decline in questions and an increase in positive sentiment from Q1 to Q2.

## Sentiment Divergence:

Banker sentiment is generally more positive than analysts, who had ~1.5 negative ratings for Credit Risk and Capital Adequacy

## Evasion hot spots:

HSBC Credit Risk evasion scores increased in Q2 vs. Q1 (from 32% to 42%).

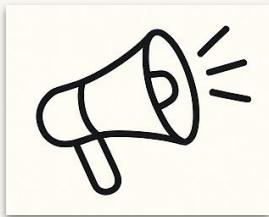
JPMorgan Operational risk evasion scores surged from 0 to 42% in Q2.

## Benchmarking:

HSBC: Cautious tone with high Capital Adequacy and Cost evasions.

JPMorgan: Governance focus with rising operational risk evasions.

# Conclusion



+



## Early Risk Detection

- Combines topic trends with shifts in sentiment around key risk areas
- **Enables supervisors to spot trends before they escalate**

## Evasion Detection

- Evasions clustering around PRA themes signal potential issues
- **Helps identify where firms may be withholding detail**

## Reliable Summarisation

- Integrates all analytical outputs
- **Delivers a short, actionable briefing for regulators**

Turning unstructured data into supervisory insight

# Q & A

Do you have any  
questions?