



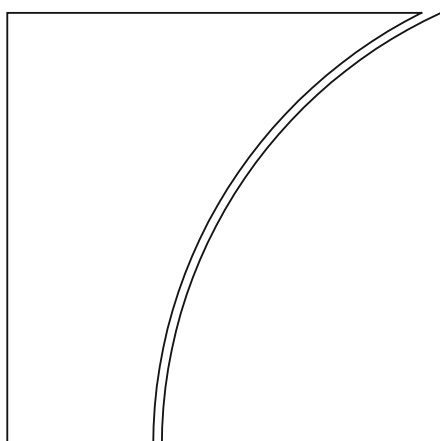
BIS Working Papers No 1299

Mapping the space of central bankers' ideas

by Taejin Park, Fernando Perez-Cruz and Hyun Song Shin

Monetary and Economic Department

October 2025



JEL classification: E52, E58, C55, C38

Keywords: central bank communication, central bank speeches, AI, topic modeling; embeddings

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in this publication are those of the authors and do not necessarily reflect the views of the BIS or its member central banks.

This publication is available on the BIS website (www.bis.org).

© *Bank for International Settlements 2025. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

Mapping the Space of Central Bankers’ Ideas

Taejin Park*

Fernando Perez-Cruz*

Hyun Song Shin*

Abstract

This paper explores the landscape of economic ideas as revealed in the machine learning embedding of a comprehensive dataset of central bank speeches. This dataset, maintained by the BIS, encompasses 19,742 speeches delivered by almost 1,000 officials from over 100 central banks over a period spanning three decades, from 1997 to 2025. As well as topic analysis of speeches at any moment in time, the evolution of the topics over time provides insights into how the focus of central bank thinking has been shaped by shifting policy challenges since 1997. Parsing the embedding both through topics and through time provides rich insights into how economic ideas have taken shape through communication practices of central banks worldwide. To demonstrate its utility, we have conducted a series of analyses that map the global landscape of monetary policy discourse. Furthermore, we construct a quantitative framework—referred to as the “*space* of central bankers’ ideas”—which uncovers institutional patterns and highlights shifts in policy approaches over time.

1 Introduction

Central bank communication is an integral part of central bank policy setting, influencing and aligning market behaviour with central bank assessments (Blinder et al. (2008)). Ben Bernanke, former chair of the Federal Reserve System once quipped that “monetary policy is 98 percent talk and 2 percent action,” underscoring the outsized role of what central bankers say in guiding expectations (Bernanke (2015)). Researchers have taken this insight to heart, increasingly leveraging natural language processing (NLP) to systematically analyse central bank communications. By converting speeches, meeting minutes, press releases and other texts into data, researchers can quantify tone, topics and signals embedded in policymakers’ words (e.g., Hansen and McMahon (2016); Gorodnichenko et al. (2023)).

The methods for analysing central bank communication have evolved alongside advances in NLP. The core approaches can be grouped into *sentiment analysis* techniques (measuring the tone or sentiment of communications) (Pang and Lee (2008)) and *topic modelling* techniques (identifying discussed themes or topics) (Vayansky and Kumar (2020)). Often these methods are combined to create richer text-derived indicators.

Early studies often relied on manual coding or simple word-count indices. A classic approach is dictionary-based sentiment analysis, using predefined word lists to measure tone. For instance, researchers frequently use the Loughran and McDonald (2011) financial lexicon to quantify positive or negative sentiment in policy texts. While such dictionaries are transparent and easy to apply, they ignore context and nuanced language. Subsequent work introduced statistical topic modelling to uncover themes in central bank communications. Latent Dirichlet Allocation (LDA) (Blei et al. (2003)) became a standard tool and has been applied to central bank speeches to identify prevalent topics. Hansen et al. (2018), for example, used topic modelling to study transparency in U.S. Federal Reserve (Fed) communications. These models, however, treat text in a “bag-of-words” manner,

*Bank for International Settlements, Basel, Switzerland. The views expressed are those of the authors and do not necessarily reflect the official positions of the Bank for International Settlements. We extend our gratitude to Rudraksh Kansal and Henry Williams for their excellent analytical and data support.

lacking consideration of word order or context. They can miss subtle linguistic signals and their static vocabulary makes it hard to capture new jargon or changing communication patterns.

In recent years, NLP in economics has embraced more sophisticated machine learning techniques. Rather than relying solely on pre-defined dictionaries or bag-of-words counts, researchers started harnessing predictive models and contextual embeddings that better capture the nuances of language. One milestone was the use of word embeddings. Early embedding methods like Word2Vec (Mikolov et al. (2013)) and GloVe (Pennington et al. (2014)) learned dense vector representations of words, allowing researchers to measure semantic similarities (e.g., “inflation” close to “prices” in vector space). Central bank researchers have trained such models on policy corpora to map the “language of central banking” in a more nuanced way (Ahrens and McMahon (2021)). Despite these advantages, these embedding models produce static vectors that assign the same representation to each word regardless of context, failing to distinguish between, for instance, a financial “bank” and a river “bank”.

To address the limitations of earlier approaches, attention has turned to transformer-based models like BERT (Devlin et al. (2019)). These models generate contextualised embeddings, capturing a word’s meaning in the specific sentence context, and can be fine-tuned for tasks such as sentiment classification. Economists have begun applying BERT and its variants to monetary policy texts. For example, Cai et al. (2021) analyse Federal Open Market Committee (FOMC) statements using BERT-based methods, and Petropoulos and Siakoulis (2021) combine textual features with XGBoost (Chen and Guestrin (2016)) to predict financial stress from central bank communications.

Gentzkow et al. (2019) surveyed “text-as-data” methods in economics and identified word embeddings as a promising frontier technology that was underutilised in economics. This expectation has indeed materialised, for example, Feldkircher et al. (2025) identify broad themes in central bank speeches, analyse their evolution and examine topic leadership among central banks. Furthermore, today’s state-of-the-art approaches focus on developing domain-specific language models tailored to the unique characteristics of central bank language. For instance, Silva et al. (2025) employ fine-tuned language models to classify individual sentences of central bank speeches across multiple dimensions—topic, communication stance, sentiment and audience. Similarly, Gambacorta et al. (2024) demonstrate the value of domain specialisation by retraining language models on a large corpus of central bank speeches and documents, finding that these specialised models outperform generic alternatives in processing central bank jargon and classifying policy stances. Beyond classification tasks, Baumgärtner and Zahner (2025) show that domain-specific embeddings trained on central bank documents can predict monetary policy shocks and reveal how deviations from inflation-targeting language affect market expectations and policy decisions.

The recent advent of chatbots based on pre-trained large language models (LLMs) like ChatGPT, Gemini or Claude has opened new frontiers for summarising and interpreting central bank communications. Researchers are exploring zero-shot or few-shot learning capabilities of such models to extract insights without extensive task-specific training. Notably, Hansen and Kazinnik (2024) ask ChatGPT to “decipher Fedspeak,” and find that GPT-4 can classify the policy stance of Fed statements with reasoning that closely matches human experts. Similarly, Alonso-Robisco and Carbó (2023) reached comparable conclusions in their study of Central Bank Digital Currency (CBDC) speeches. This suggests LLMs can capture subtle signals in various central bank topics that previous algorithms might miss.

Beyond classification and summarisation, machine-learning techniques enable new ways to visualise and compare central bank communications. By encoding speeches into high-dimensional embeddings, one can measure similarities between communications and track how topics cluster. Modern topic modelling approaches like BERTopic (Grootendorst (2022)) leverage such embeddings to discover latent themes. BERTopic uses LLM-generated embeddings to group documents into clusters, then identifies each cluster’s key terms using a class-based TF-IDF approach (Salton and Buckley (1988)). This contextualised approach yields more coherent topics than traditional methods because the grouping is based on contextual semantic similarity rather than just word co-occurrence. Such transformer-based topic modelling can flexibly capture new relevant vocabulary (e.g., “quantitative easing” or “cryptoassets”) and better separate closely related themes, thanks to the contextual embeddings.

Furthermore, once speeches are represented as embeddings, researchers can project them into lower-dimensional maps for intuitive visualisation, representing context-dependent relationships

that align more closely with human knowledge (Grand et al. (2022)). Techniques like t-distributed stochastic neighbor embedding (t-SNE, van der Maaten and Hinton (2008)) and Uniform Manifold Approximation and Projection (UMAP, McInnes et al. (2020)) apply nonlinear dimensionality-reduction methods tailored for visualising high-dimensional data by preserving local relationships.

This paper demonstrates how to apply NLP methods to analyse the global landscape of monetary policy discourse. By leveraging LLMs, contextual embeddings, topic modelling and dimensionality reduction techniques, we construct a quantitative "space of central bankers' ideas" that maps this landscape. Our approach showcases how modern NLP can reveal institutional communication patterns, track policy evolution and identify emerging themes in central banking. The resulting visualisations offer an intuitive template for exploring the multidimensional nature of central bank communications, enhancing both academic research and practical understanding of monetary policy discourse. This work serves as a methodological blueprint for applying cutting-edge text analysis to central bank communications.

2 Data construction and description

Research progress in analysing central bank speeches has been constrained by the historically limited availability and fragmented nature of comprehensive, structured and machine-readable datasets. Recent efforts have begun to address this gap—the European Central Bank (ECB) now provides speeches in a machine-readable format. At the Bank for International Settlements (BIS), we have compiled a dataset comprising speeches from over 100 central banks across advanced and emerging market economies, providing comprehensive coverage of the global monetary policy landscape. The dataset covers speeches from 1997 onwards, capturing significant periods of recent economic history including financial crises, policy regime changes and the emergence of unconventional monetary policies. We have supplemented the raw speech data with essential metadata and additional features to facilitate systematic NLP analysis and cross-institutional comparisons.

Specifically, we augment the corpus with comprehensive speaker metadata, including institutional affiliations and geographical locations, thereby facilitating both institution-specific analyses and cross-economy comparative investigations. Furthermore, we incorporate hierarchical speaker position classifications, enabling researchers to account for the heterogeneous impact of communications contingent upon the speaker's organisational authority.² To enhance analytical accessibility, we supplement the dataset with AI-generated standardised summaries, topic representations and pre-computed high-dimensional embeddings—which we use in our applications in this paper.

2.1 Data collection and processing

The dataset used in our paper comprises 19,742 unique English speech transcripts delivered by governors, deputy governors and senior officials over 100 central banks globally between January 1997 and March 2025.

Building upon the BIS dataset foundation, we implemented metadata enrichment. To extract comprehensive information about speakers, we employed a multi-faceted approach. First, we compiled authoritative lists of senior central bank officials and their tenure periods for major central banks (including, but not limited to, the Fed, ECB, Bank of Japan (BoJ), Bank of England (BoE), Reserve Bank of India (RBI), South African Reserve Bank, Banco Central do Brasil) using official institutional websites, enabling precise attribution of speeches by high-ranking officials. Second, we systematically processed metadata from the original source, which typically included speech titles, brief descriptions and full transcripts containing speaker information. These elements were individually submitted to an LLM (OpenAI's GPT-4o, accessed via API in May 2025) to extract speaker details in a structured format. Third, we leveraged the parametric knowledge of the language model by providing speaker names and speech dates to retrieve corresponding institutional affiliations and positions at time of delivery. We subsequently consolidated these three independent data collection approaches to ensure robustness. In approximately 97% of cases, the approaches yielded consistent information. For the remaining 3% where discrepancies emerged, we conducted manual verification and correction procedures.

²Ahrens et al. (2025) found that speaker position significantly influences market reactions to central bank communications.

2.2 Descriptive statistics

The dataset includes, for each speech, the URL, title, description, date, full text, standardised summary, speaker name, position, institutional affiliation, ISO2-coded location, topic representations and precomputed embeddings.

The corpus shows increasing density after 2008, reflecting the enhanced emphasis on central bank communication following the Great Financial Crisis (Figure 1a). This temporal coverage enables analysis of communication strategy evolution across distinct economic regimes, including conventional and unconventional monetary policy implementation periods, financial crises and varying inflation environments, facilitating comparative analysis of institutional responses through communication channels.

Among communications from over 100 central banks globally, the ECB and the Fed collectively constitute approximately 25% of the corpus (Figure 1b). For analytical coherence, the twelve Federal Reserve Districts are represented as a unified institution, while ECB speeches exclude communications from officials when representing their national central banks rather than the supranational institution.

The corpus includes communications from about 1,000 unique central bank officials. Our approach accommodates cases where individuals held multiple roles throughout their careers. Speech distribution exhibits significant concentration among institutional leadership, with former ECB President Jean-Claude Trichet emerging as the most prolific communicator (appearing in 478 speeches in our corpus), followed by his successor Mario Draghi (337), Amando M. Tetangco, Jr. (272) of the Bangko Sentral ng Pilipinas (BSP) and Ben S. Bernanke (253) of the Fed, as shown in Figure 1c. These most frequent communicators all held leadership positions during the post-2008 great financial crisis era, highlighting how this watershed event transformed central bank communication into a critical policy instrument.

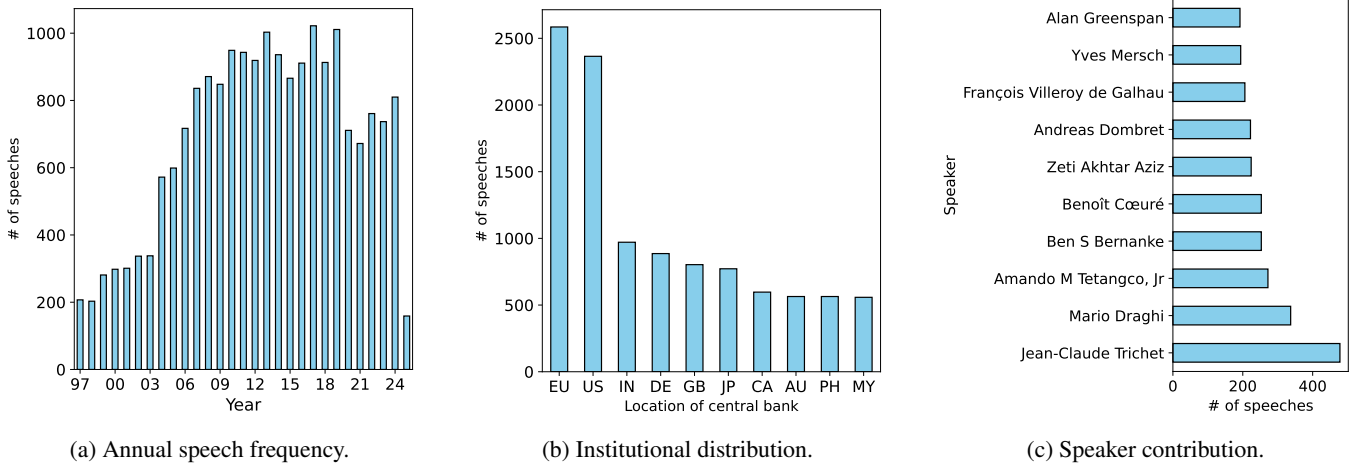


Figure 1: Central bank communication patterns.

3 Methodology

This section provides methodological details for the analyses. The overall workflow for embedding generation, topic modelling and visualisation is illustrated in Figure 2.

3.1 Embedding process

Raw speech transcripts exhibited considerable variation in format, style, length and structure. To mitigate the risk that analytical outcomes might reflect these presentational inconsistencies rather than substantive content differences across temporal and institutional dimensions, we implemented a standardisation process. Specifically, we employed GPT-4o to transform each speech into a

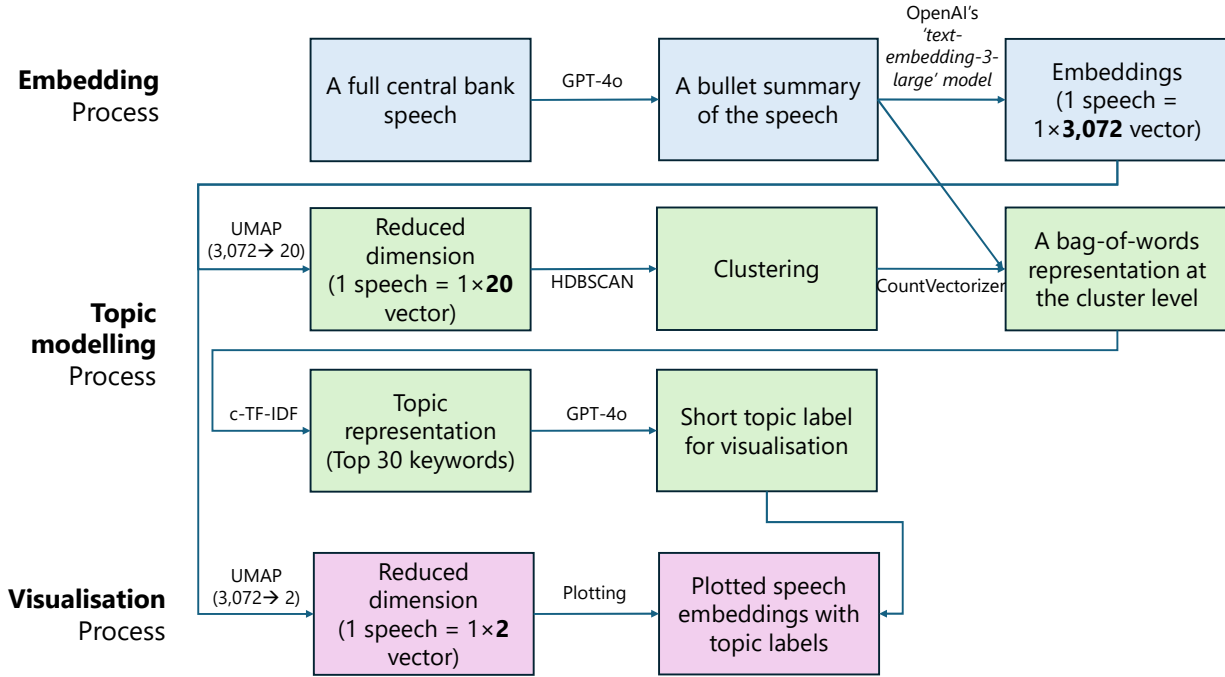


Figure 2: Workflow for embedding, topic modelling and visualisation of central bank communications.

standardised format consisting of concise bullet points.³ This procedure removes non-essential elements (e.g., greetings, acknowledgments and disclaimers), eliminates structural irregularities and normalises content presentation. After standardisation, we converted each speech summary into a high-dimensional vector representation using OpenAI's 'text-embedding-3-large' model.⁴ This transformer-based model generates 3,072-dimensional embeddings that capture the semantic content of each speech. Each vector represents a semantic "fingerprint" of a speech, encoding its topical content, tone and policy stance in a format suitable for computational analysis.

3.2 Topic modelling with BERTopic

To identify thematic patterns in the speech corpus without imposing predetermined categories, we employed the BERTopic framework. The process involved three key steps: dimensionality reduction, clustering and topic representation.

The initial 3,072-dimensional embeddings, while rich in semantic information, are excessively high-dimensional for efficient clustering. We therefore applied UMAP to reduce the embeddings from 3,072 to 20 dimensions. UMAP preserves local semantic relationships between speeches while making the data amenable to clustering.

³See Appendix C.1 for the used prompt. We conducted this process for the entire dataset in June 2024 and again in May 2025 to incorporate updates.

⁴Accessed on 27–28 August 2024 for data up to March 2024, and on 7 May 2025 for the latest updates.

UMAP is a dimensionality reduction technique designed to effectively preserve both local and global structures of data. Unlike linear methods such as principal component analysis, UMAP captures non-linear relationships in high-dimensional data by constructing a topological representation of the data manifold. It first builds a weighted nearest-neighbor graph in high-dimensional space, then optimises a lower-dimensional embedding to preserve these relationships. This makes UMAP suitable for our application, as it maintains semantic similarities between speeches, ensuring communications on similar topics remain closely positioned in the reduced-dimensional space. Additionally, it retains broader structural patterns that reflect institutional and regional groupings. Reducing the dimensionality from 3,072 to 20 dimensions preserves essential complexity for accurate clustering while ensuring computational efficiency.

With the reduced-dimension embeddings, we applied the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN, McInnes et al. (2017)) algorithm to identify natural groupings of related speeches. HDBSCAN has several advantages for our application: it does not require pre-specifying the number of clusters, flexibly accommodates varying cluster shapes and sizes and labels outliers as noise rather than forcing them into arbitrary clusters. This approach allowed topics to emerge organically from the data rather than imposing an arbitrary topic structure, making it well-suited for exploratory analysis of a diverse corpus spanning multiple institutions and time periods.

To characterise the content of each identified topic, we first applied CountVectorizer to convert speech text into a bag-of-words representation at the cluster level. We also removed stop words for more interpretable representations. Next, we applied class-based TF-IDF to enhance topic representation by emphasising the most distinctive words within each cluster. As a result, the topic of each cluster is represented as a vector of words (Table 1).

3.3 Visualisation & analysis process

To facilitate intuitive interpretation and comparative analysis of central bank communications, we projected the initial 3,072-dimensional speech embeddings onto two-dimensional maps using UMAP.

This visualisation approach enables the construction of a conceptual landscape, a visual “space of central bankers’ ideas”, where proximity between points represents semantic similarity between speeches. In the resulting visualisation, each point represents a single speech, colour-coded according to its assigned topic from the topic modelling process described in Section 3.2. For visual clarity, the generic topic (-1) that is not closely assigned to any clusters is coloured light grey.

It is important to acknowledge that topic modelling operates in a higher-dimensional space (20 dimensions), whereas this visualisation is based on a two-dimensional projection. As such, the visualised clusters may not fully capture the structure of the topic assignments. Some topic boundaries appear blurred or overlapping in 2D, even though the modelling aims to achieve clearer separation in the high-dimensional space. This trade-off in dimensionality reduction enhances interpretability while striving to preserve the most meaningful semantic relationships from the original space.

Furthermore, we incorporated the temporal dimension into this semantic space to analyse evolutionary trends in central bank communications. This integration enabled us to track how communication priorities shifted in response to changing economic conditions over time.

Our analytical approach followed an iterative “zooming-in” strategy across three levels of granularity:

1. **Global-level analysis:** We first examined the entire speech corpus to identify overarching themes in monetary policy discourse over the nearly three-decade span of our dataset. This global view revealed regional clustering patterns and highlighted the evolution of central bank communication priorities.
2. **Institutional-level analysis:** We narrowed our focus to a specific central bank, the ECB. This allowed us to identify institutional “communication signatures.”
3. **Topic-level analysis:** At a finer granularity, we performed targeted analysis of ECB monetary policy communications, dissecting subtle thematic distinctions within the core policy domain. This revealed how communication on core topics like inflation evolved across different economic environments and policy regime.

4 Results

We now turn to the empirical findings from our analysis of central bank communications. For these analyses, we focus on 40 major central banks spanning diverse economic regions and monetary policy frameworks.⁵ We begin with the global communication landscape, where the full power of our approach becomes evident through the emergence of clear regional clusters, institutional signatures, and thematic evolution over time in an unsupervised way.

4.1 Global communication landscape analysis

The dataset facilitates comprehensive mapping of the global central banking discourse landscape, revealing structural patterns that would be difficult to detect through traditional analysis methods.

When visualising the entire corpus of central bank speeches in a two-dimensional embedding space, several clear patterns emerge (Figure 3). First and most prominently, we observe strong regional clustering of topics, with speeches clustering primarily by country and region. Beyond individual central bank clusters, we observe distinct groupings for Asian central banks (Topic 12) and those focused on Islamic finance (Topic 19), indicating shared policy priorities and communication patterns among central banks operating in similar economic contexts. Furthermore, central banks with certain shared characteristics are located in proximity within this vector space—Asian central bank clusters concentrate in one region, while those of advanced economies concentrate in another.

Two interpretations of this geographical clustering pattern are possible. On one hand, central banks globally may discuss similar themes due to their comparable objectives and mandates, with geographical markers serving primarily as distinctive labels that the topic model has identified. The observed clustering might therefore reflect differences in regional terminology or institutional phrasing rather than fundamental differences in policy content. On the other hand, despite increasing coordination among central banks, communication patterns may remain significantly influenced by regional economic conditions, institutional contexts and policy frameworks, resulting in genuinely different topical emphases across regions.

Large central banks like the Fed (Topics 1, 27, 35) and the ECB (Topics 0, 15, 16, 29, 32, 47) form multiple topic clusters within the embedding space. The ECB, in particular, forms six distinct topic clusters. This likely reflects institutional characteristics unique to the ECB, such as an expansive communication strategy, multinational governance structure and a mandate that explicitly extends beyond the objective of price stability to support the general economic policies in its member states. This multiplicity of communication clusters from major central banks motivates our subsequent institutional-level analysis, where we focus specifically on a large central bank to analyse communication patterns in greater depth.

Deeper analysis of the topic representations reveals distinctive policy emphases within each national or regional cluster, extending beyond traditional central banking themes. These distinctive communication priorities appear to reflect each central bank's specific policy challenges and institutional mandates. As shown in Table 1, the ECB cluster, for instance, displays emphasis on fiscal policy, structural reforms and integration—themes that reflect the ECB's unique position within the Eurozone architecture. In contrast, the BoJ cluster exhibits a concentrated focus on monetary easing strategies and Quantitative and Qualitative Monetary Easing (QQE) programmes, reflecting Japan's prolonged battle with deflationary pressures and economic stagnation.

Central banks from smaller open economies with distinctive policy frameworks exhibit specialised communication patterns reflecting their unique economic contexts. For instance, the central banks of Denmark and Switzerland, which operate within economies where exchange rate stability plays a central role in monetary policy implementation, show a higher emphasis on currency-related topics in their communications. These findings suggest that while central banks globally share certain core objectives such as price stability, their communications reflect specialised policy frameworks adapted to distinct regional economic structure and challenges.

Within this regional clustering, we identify a central hub in the embedding space comprising topics that require significant international cooperation. These include discussions of banking regulations, the global economy and monetary and financial stability (Topics 5, 7, 21). The centrality of these

⁵See Appendix A for the full list of central banks included in our analysis.

topics in the space may reflect their importance as points of convergence in central bank discourse across regions. This finding aligns with the increasing coordination among central banks following the 2008 financial crisis and subsequent efforts to harmonise regulatory frameworks.

Notably, several emerging topics form isolated clusters at the periphery of the embedding space. Climate change (Topic 18) represents the most prominent example, forming a distinct cluster as climate-related discussions have gained prominence in central bank communications. Similar patterns appear for other emerging topics such as benchmark reforms (Topic 37 covering Libor transition, FX Global Code, etc), which occupy distinct regions.

As discussed in the previous section, the use of different dimensionalities for topic modelling and visualisation may result in some seemingly outlying speeches. These outliers do not indicate noise but rather reflect specific topical discussions (topical cluster) relevant to the unique economic contexts of those countries (country or regional cluster).

Table 1: Selected examples of topic representations

Indicative label	Top keywords
ECB	euro, ecb , area, euro area, european, monetary, economic, stability, policy, union , fiscal , growth, reforms , monetary policy, price, structural , price stability, inflation, countries, financial, policies, eu, structural reforms , europe, rates, market, integration, ...
JP	japan , economic, bank japan, prices, growth, easing , corporate, bank, economy, monetary easing , monetary, recovery, price, japan economy, investment, demand, private, qqe , ...
DK	denmark , danish, nationalbank, danmarks, danmarks nationalbank, fixed exchange rate , ...
CH	snb, swiss, switzerland , franc, swiss franc , ...

Note. Keywords are ordered by probability (importance), with higher probability terms listed first. The **indicative labels** in the left column are automatically assigned by GPT-4o based on the top keywords for interpretive clarity. The full list of topic representations is available in Appendix B.

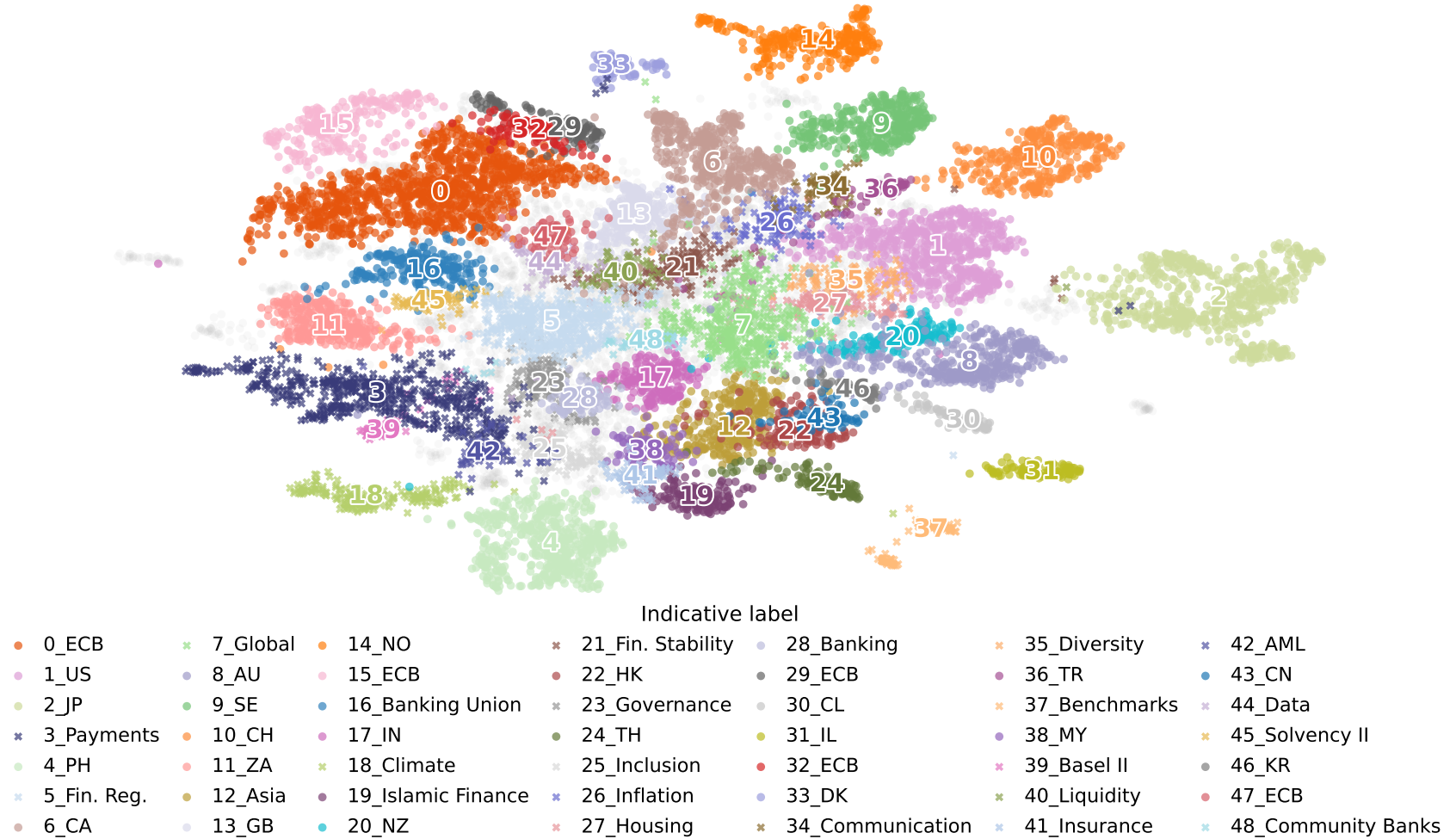


Figure 3: Global topic distribution of central bank communications (1997-2025).

Note: For conciseness, we leveraged GPT-4o to automatically assign topic labels based on the predominant themes identified. The specific prompt used for this purpose is detailed in Appendix C.2. Lower topic numbers correspond to larger topic clusters. For enhanced readability, the geographical clusters are represented as dots while topical clusters are marked with crosses. Speeches not belonging to any topic cluster are shown in light grey. The topic modelling used an n-gram range of (1,4) with a minimum cluster size of 50 speeches. Euclidean distance was used for UMAP dimensionality reduction. The full list of topic representations is available in Appendix B.

To further understand who emphasises which topics, we disaggregate key topical clusters by the institutional affiliation of the speakers (Figure 4). As expected, the largest central banks, such as the Fed and the ECB, account for the largest share of speeches on globally relevant themes like the global economy and financial stability. In the domain of financial regulations, central banks with large financial sectors, such as the Fed and the BoE, feature prominently. Interestingly, the BoE delivers a disproportionately high number of speeches on Solvency II, an EU-wide insurance regulatory framework, even more so than the ECB. This discrepancy reflects institutional differences: the BoE, through the Prudential Regulation Authority, is responsible for implementing and supervising insurance regulations in its economy, while the ECB does not oversee insurance firms. Emerging themes such as climate change and financial inclusion also reveal institution-specific communication priorities. Financial inclusion is discussed primarily by the RBI and Bank Negara Malaysia, reflecting their unique explicit developmental mandates.

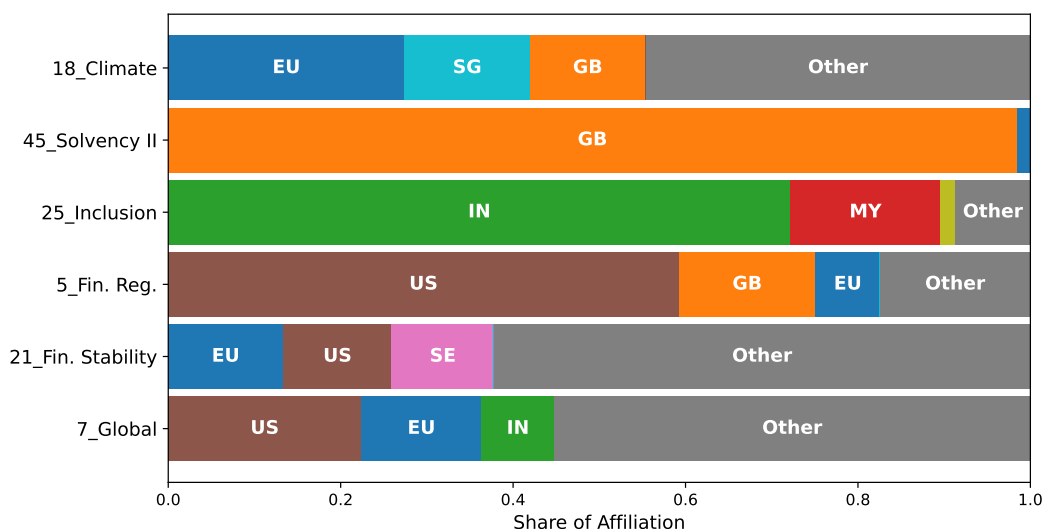


Figure 4: Topic by economy.

Note: Each bar is ordered by the share of speeches, with larger shares displayed on the left.

Additionally, we incorporate the time dimension to reveal how these patterns evolve. To capture this temporal evolution, we projected the speech embeddings into a three-dimensional space, where the third dimension represents time (Figure 5), with snapshots over three periods (Figure 6). This visualisation reveals several intriguing patterns in the temporal distribution of central bank communication topics.

Traditional focus of central banks like inflation expectation discussions (Topic 26) appears consistently throughout our sample period, reflecting the enduring centrality of this mandate to central banking. This persistence underscores that regardless of changing economic conditions or emerging challenges, inflation concerns remain a fundamental and time-invariant component of central bank discourse.

In contrast, some other topics show distinct temporal patterns of emergence. For instance, data (or statistics) receives minimal attention until April 2004, when Jean-Claude Trichet delivered a speech emphasising the critical role of data in ECB policy-making.⁶ Following this inflection point, we observe that statistical topics became a recurring theme in central bank communications, illustrating how individual speeches can sometimes catalyse broader discussions within the central bank community.

The three-dimensional visualisation also clearly illustrates the rise and decline of specific policy priorities. Climate change discussions show a pronounced increase in recent years, reflecting growing central bank acknowledgment of climate-related financial risks. Conversely, the Basel II implementation topic demonstrates temporary prominence during the development and implementation phases before gradually declining as the framework became established.

⁶<https://www.ecb.europa.eu/press/key/date/2004/html/sp040422.en.html>

More interesting are topics that maintain consistent presence but undergo substantive evolution in content. Payment systems discussions exemplify this pattern, appearing consistently throughout our sample but showing dramatic intensification in recent years with the emergence of fintech innovations, digital forms of cash and blockchain technologies. This pattern illustrates how central banks continuously reinterpret and expand traditional domains of responsibility in response to technological and market developments.

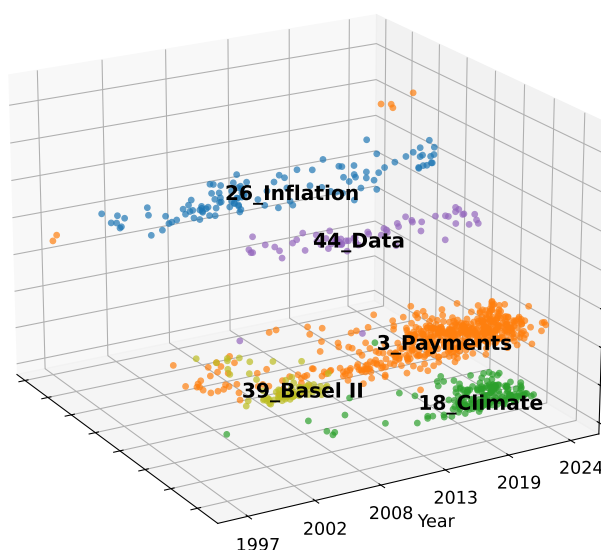


Figure 5: Temporal evolution of selected topics.

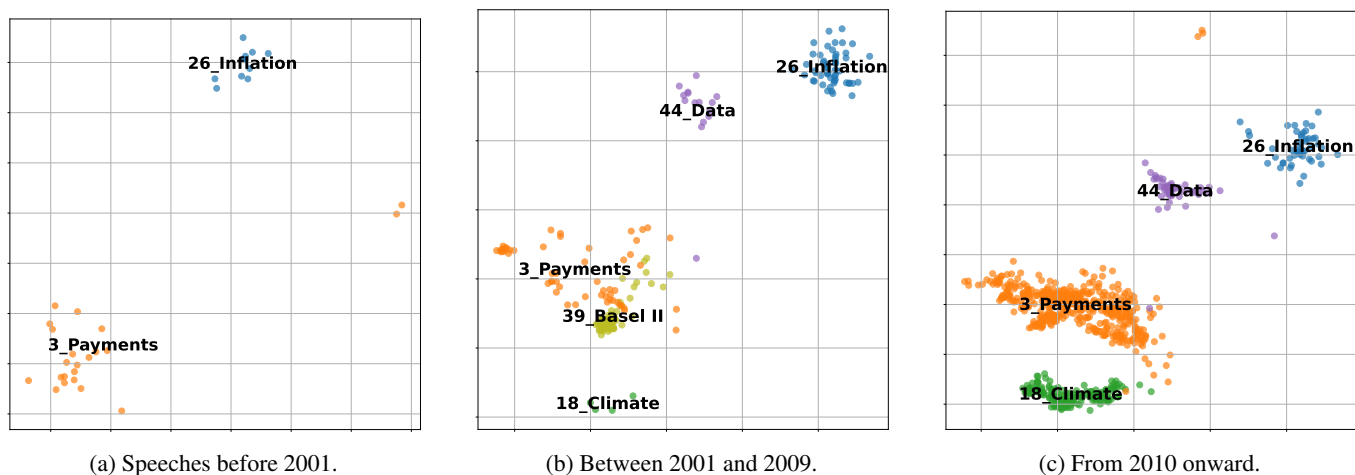


Figure 6: Temporal snapshots of selected topics across three periods.

4.2 Deep dive 1: Analysis of ECB communications

We now turn to institution-specific analysis, examining ECB communications to demonstrate how our approach reveals a central bank's core functions and policy priorities. For this analysis, we conduct a separate topic modelling exercise focused exclusively on ECB speeches. Rather than simply extracting ECB clusters from our global analysis, we reapply our methodology—topic modelling

and UMAP dimension reduction for visualisation—to the subset of ECB speeches. This targeted approach enables finer-grained topic detection and uncovers institutional patterns that remain hidden in the broader global analysis.

Figure 7 reveals that ECB communications organise into three primary thematic regions: monetary policy (Topics 0, 2, 5, 7, 13, 14), financial stability (Topics 1, 6, 18) and market integration specific to the currency union (Topics 3, 10, 19). Additionally, the ECB has progressively incorporated discussions of structural challenges, including structural reforms (Topic 17), household inequality and wealth (Topic 15) and climate change (Topic 11), into its communications portfolio, reflecting its secondary mandate to support broader EU economic policies, as established in the Treaty on the Functioning of the European Union, Article 127(1).

Upon closer examination of the monetary policy-related topic clusters, we observe that these topics are frequently distinguished by temporal characteristics, with topic representations often including specific time periods. This temporal specificity suggests that ECB monetary policy communications are strongly influenced by the prevailing economic context and policy regime. To investigate this phenomenon further, we focused on the core monetary policy function by selecting specific clusters (Topics 0, 2, 4, 5, 7, 13, 14) and conducting a more granular topic modelling and visualisation exercise within this subset.

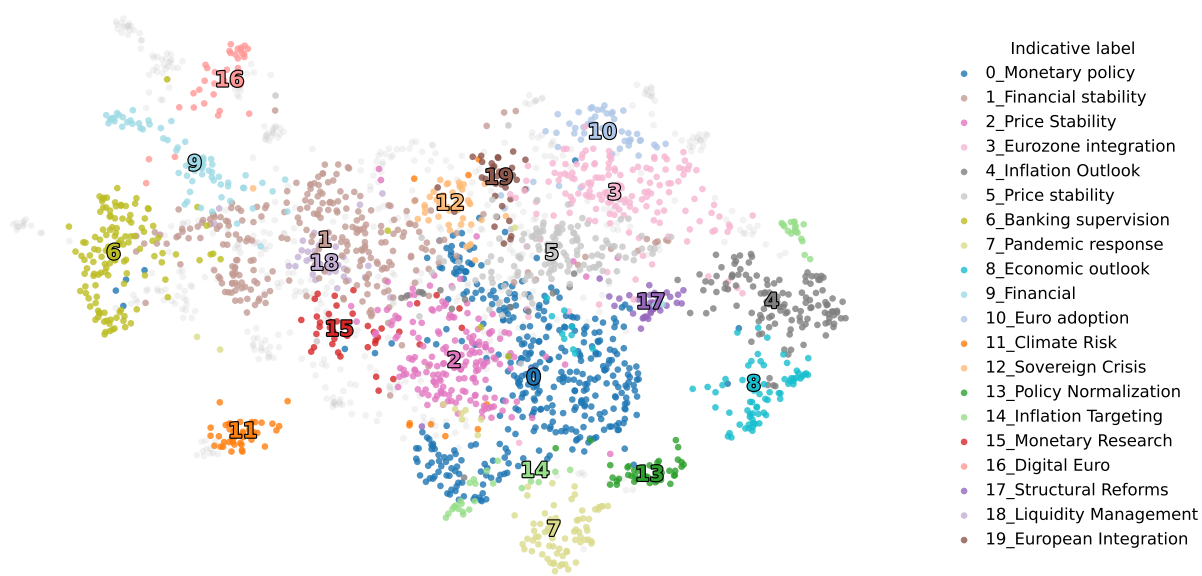


Figure 7: Semantic map of ECB communications

Note: We leveraged GPT-4o to automatically assign topic labels based on the predominant themes identified. The specific prompt used for this purpose is detailed in Appendix C.3. Lower topic numbers correspond to larger clusters. Speeches not belonging to any topic cluster are shown in light grey. The topic modelling used an n-gram range of (1,2) with a minimum cluster size of 30 speeches. Euclidean distance was used for UMAP dimensionality reduction.

4.3 Deep dive 2: In-depth analysis of ECB’s core communications

To examine these monetary policy communications at an even finer granularity, we reapply our methodology, but now to a more focused subset of monetary policy-related clusters identified in the previous ECB analysis and conduct a third iteration of topic modelling and visualisation. Our analysis of two-dimensional embeddings supplemented with temporal data confirms that ECB communication topics largely cluster chronologically (Figure 8). While the largest monetary policy cluster (Topic 0, light grey cluster) maintains a persistent presence throughout the sample period, other thematic clusters demonstrate pronounced temporal concentration, suggesting an adaptive communication strategy that evolves in response to prevailing economic conditions.

Particularly noteworthy are the brown and green clusters, which exhibit remarkable temporal specificity in recent years. The brown cluster can be characterised by pandemic-related communications, representing an expected rhetorical shift in response to an exogenous economic shock of unprecedented magnitude that necessitated rapid policy adaptation and novel crisis communication strategies. More intriguing is the green cluster, which encompasses inflation-related communications, despite inflation being a fundamental component of the ECB’s primary mandate. When contextualised within the macroeconomic environment, this temporal concentration becomes more comprehensible. The 2022–2023 period presented extraordinary monetary policy challenges characterised by high inflation rates, levels unseen in the euro area since the institution’s establishment. This exceptional inflationary regime required an equally exceptional communication response. Similarly, other chronological clusters reflect particular policy regimes: the initial establishment period of the ECB (blue), the pre-crisis period of conventional monetary policy (orange) and the quantitative easing and unconventional policy period following the sovereign debt crisis (purple).

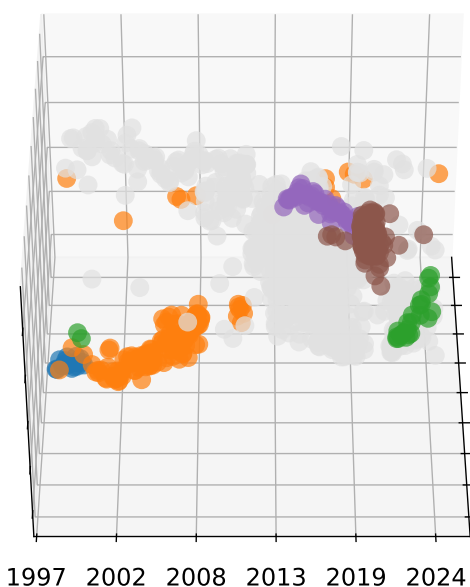


Figure 8: Temporal evolution of ECB monetary policy communication.

Note: Topic clusters correspond to distinct temporal policy regimes: green (inflation-focus), brown (pandemic response), purple (unconventional monetary policies), orange (conventional monetary policies) and blue (institutional establishment phase). For visual clarity, the largest topic cluster, which encompasses general monetary policy terminology, is displayed in light grey. The topic modelling used an n-gram range of (1,2) with a minimum cluster size of 10 speeches. Euclidean distance was used for UMAP dimensionality reduction.

4.4 Validation through sentiment analysis

To validate that the temporal topic clusters identified in Figure 8 genuinely correspond to macroeconomic conditions rather than merely capturing superficial linguistic patterns, we conducted a complementary sentiment analysis of the monetary policy stances within each cluster. This analysis provides an independent verification that our topic modelling approach captures substantive policy content.

We employed the GPT-4o model to classify the monetary policy sentiment of speeches within each temporal cluster as hawkish, neutral or dovish.⁷ To account for the probabilistic nature of LLM classifications and capture sentiment intensity, we repeated each classification five times and computed the average score on a scale from -1 (most dovish) to +1 (most hawkish). This approach yields a nuanced measure of policy stance that goes beyond binary classifications.

⁷See Appendix C.4 for the prompt used. This classification process was conducted in June 2025.

Figure 9 reveals how topic clusters align with their expected monetary policy orientations over time. The unconventional monetary policy cluster (purple) maintains consistently dovish sentiment scores throughout its period, reflecting the sustained accommodative stance required during the post-crisis era. Similarly, the pandemic response cluster (brown) exhibits uniformly dovish sentiment, confirming that these clusters capture genuine policy regime shifts requiring substantial monetary accommodation.

Particularly revealing is the pattern within the recent high-inflation cluster (green). The sentiment analysis shows this cluster bifurcates into an initial hawkish phase followed by a dovish turn as inflationary pressures began to moderate. This temporal evolution within a single topic cluster demonstrates that our methodology captures not just discrete policy regimes but also dynamic transitions within extraordinary economic episodes. The high-inflation period emerges as a unified topic because it represents a coherent policy challenge, the management of unprecedented inflation, despite encompassing both the tightening and subsequent easing phases of the policy response. The conventional monetary policy regime (orange) exhibits similar cyclical patterns, with alternating dovish and hawkish sentiment that tracks the ebb and flow of inflationary pressures over business cycles.

These sentiment patterns provide additional validation that our topic modelling approach identifies meaningful policy regimes that align with economic fundamentals and actual policy stances, rather than merely detecting linguistic similarities.

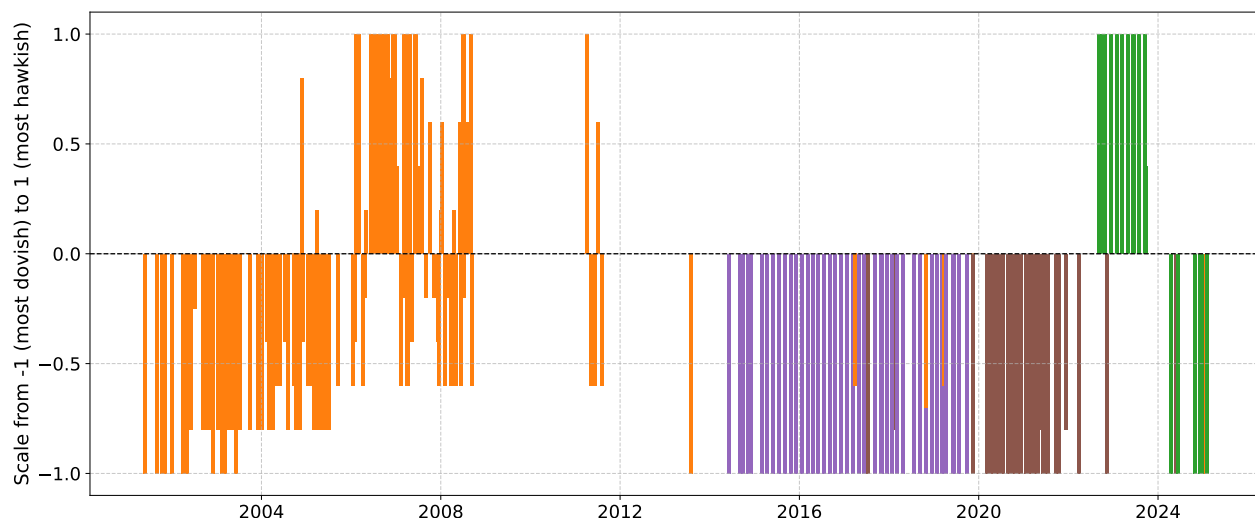


Figure 9: ECB monetary policy sentiment validation of temporal topic clusters.

Note: The colours correspond to the temporal topic clusters identified in Figure 8: green (inflation-focus), brown (pandemic response), purple (unconventional monetary policies) and orange (conventional monetary policies). This graph displays the average monetary policy sentiment classification for speeches within each cluster, where sentiment scores range from -1 (most dovish) to +1 (most hawkish). Each speech was classified five times using ChatGPT to account for probabilistic variation, with scores averaged to capture sentiment intensity.

5 Conclusion

This research demonstrates how advanced NLP techniques, particularly LLMs, provide powerful new tools for analysing the global landscape of central banking communications at scale. By leveraging embedding-based approaches and dimensionality reduction, we have constructed a quantitative "space of central bankers' ideas" that reveals institutional patterns, policy priorities, and evolutionary trends across nearly three decades of monetary policy discourse.

Our methodology enables efficient analysis of similarities and differences between institutions, identification of common and unique challenges and tracking of evolving communication strategies in ways that were previously limited to qualitative assessment. The visualisations created through our approach offer an intuitive interface for exploring the multidimensional nature of central bank

communications, enhancing transparency and understanding of monetary policy discourse across global, institutional and topical levels.

Notably, we found that speech embeddings capture not only topical content but also incorporate economic contexts and policy sentiments. This research opens several promising avenues for future work. The methodological framework could be extended to create real-time monitoring systems for central bank communication shifts. Additionally, our database of embeddings enables integration with traditional econometric analysis, allowing researchers to incorporate textual data alongside economic indicators for more comprehensive modelling. As central bank communication continues to evolve as a key policy tool, quantitative methods like those presented here will play an increasingly important role in understanding and evaluating the economic policy context.

References

- Ahrens, M., Erdemlioglu, D., McMahon, M., Neely, C. J., and Yang, X. (2025). Mind your language: Market responses to central bank speeches. *Journal of Econometrics*, 247(105921).
- Ahrens, M. and McMahon, M. (2021). Extracting economic signals from central bank speeches. In *Proceedings of the Third Workshop on Economics and Natural Language Processing*, pages 93–114.
- Alonso-Robisco, A. and Carbó, J. M. (2023). Analysis of CBDC narrative by central banks using large language models. *Finance Research Letters*, 58(104643).
- Baumgärtner, M. and Zahner, J. (2025). Whatever it takes to understand a central banker - embedding their words using neural networks. *Journal of International Economics*, 157(104101).
- Bernanke, B. S. (2015). Inaugurating a new blog. <https://www.brookings.edu/articles/inaugurating-a-new-blog/>. Ben Bernanke’s Blog, March 30.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.
- Blinder, A. S., Ehrmann, M., Fratzscher, M., De Haan, J., and Jansen, D.-J. (2008). Central bank communication and monetary policy: A survey of theory and evidence. *Journal of Economic Literature*, 46(4):910–945.
- Cai, Y., Camara, S., and Capel, N. (2021). It’s not always about the money, sometimes it’s about sending a message: Evidence of informational content in monetary policy announcements. *arXiv:2111.06365 [econ.GN]*.
- Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16, page 785–794. ACM.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Burstein, J., Doran, C., and Solorio, T., editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Feldkircher, M., Koráb, P., and Teliha, V. (2025). What do central bankers talk about? evidence from the bis archive. CAMA Working Paper 35/2025, Centre for Applied Macroeconomic Analysis (CAMA).
- Gambacorta, L., Kwon, B., Park, T., Patelli, P., and Zhu, S. (2024). CB-LMs: Language models for central banking. *BIS Working Papers*, 1215.
- Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–74.
- Gorodnichenko, Y., Pham, T., and Talavera, O. (2023). The voice of monetary policy. *American Economic Review*, 113(2):548–84.
- Grand, G., Blank, I. A., Pereira, F., et al. (2022). Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*, 6:975–987.
- Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv:2203.05794*.
- Hansen, A. L. and Kazinnik, S. (2024). Can ChatGPT decipher FedSpeak? Available at SSRN: <https://ssrn.com/abstract=4399406> or <http://dx.doi.org/10.2139/ssrn.4399406>.
- Hansen, S. and McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99:S114–S133.

- Hansen, S., McMahon, M., and Prat, A. (2018). Transparency and deliberation within the FOMC: A computational linguistics approach. *The Quarterly Journal of Economics*, 133(2):801–870.
- Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65.
- McInnes, L., Healy, J., and Astels, S. (2017). HDBSCAN: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205.
- McInnes, L., Healy, J., and Melville, J. (2020). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv:1802.03426 [stat.ML]*.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv:1301.3781 [cs.CL]*.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.
- Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. Association for Computational Linguistics.
- Petropoulos, A. and Siakoulis, V. (2021). Can central bank speeches predict financial market turbulence? Evidence from an adaptive NLP sentiment index analysis using XGBoost machine learning technique. *Central Bank Review*, 21(4):141–153.
- Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523.
- Silva, T. C., Moriya, K., and Veyrune, R. (2025). From Text to Quantified Insights. *IMF Working Papers*, 109.
- van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605.
- Vayansky, I. and Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, 94(101582).

Appendix

A Central banks included in analysis

Code	Central bank	Code	Central bank
AE	Central Bank of the United Arab Emirates	KR	Bank of Korea
AR	Central Bank of Argentina	KW	Central Bank of Kuwait
AU	Reserve Bank of Australia	MA	Bank Al-Maghrib
BR	Central Bank of Brazil	MX	Bank of Mexico
CA	Bank of Canada	MY	Bank Negara Malaysia
CH	Swiss National Bank	NO	Norges Bank
CL	Central Bank of Chile	NZ	Reserve Bank of New Zealand
CN	People's Bank of China	PE	Central Reserve Bank of Peru
CO	Central Bank of Colombia	PH	Bangko Sentral ng Pilipinas
CZ	Czech National Bank	PL	Narodowy Bank Polski
DK	Danmarks Nationalbank	RO	National Bank of Romania
DZ	Bank of Algeria	RU	Central Bank of the Russian Federation
EU	European Central Bank	SA	Saudi Central Bank
GB	Bank of England	SE	Sveriges Riksbank
HK	Hong Kong Monetary Authority	SG	Monetary Authority of Singapore
HU	Magyar Nemzeti Bank	TH	Bank of Thailand
ID	Bank Indonesia	TR	Central Bank of the Republic of Türkiye
IL	Bank of Israel	US	Federal Reserve System
IN	Reserve Bank of India	VN	State Bank of Vietnam
JP	Bank of Japan	ZA	South African Reserve Bank

B Topic representations of the global central bank communication

Topic	Label	Top 30 keywords
-1	Generic	financial, policy, economic, market, banks, stability, monetary, central, risk, bank, crisis, inflation, credit, growth, banking, global, capital, monetary policy, markets, financial stability, risks, term, need, liquidity, significant, including, management, sector, reserve, rates
0	ECB	euro, ecb, area, euro area, european, monetary, economic, stability, policy, union, fiscal, growth, reforms, monetary policy, price, structural, price stability, inflation, countries, financial, policies, eu, structural reforms, europe, rates, market, integration, term, measures, single
1	US	federal, inflation, economic, policy, federal reserve, labor, reserve, growth, fomic, rate, market, employment, labor market, unemployment, monetary, monetary policy, rates, federal funds, term, spending, funds rate, conditions, federal funds rate, fed, economy, low, remains, recovery, funds, percent
2	JP	japan, economic, bank japan, prices, growth, easing, corporate, bank, economy, monetary easing, monetary, recovery, price, japan economy, investment, demand, private, qqe, consumption, boj, financial, firms, rates, policy, inflation, expected, term, business, conditions, target
3	Payments	payment, payments, digital, fintech, innovation, systems, financial, bank, money, services, payment systems, cash, central, banks, settlement, electronic, retail, new, technology, like, regulatory, data, efficiency, cbdc, risks, sepa, transactions, security, central bank, cross
4	PH	bsp, financial, philippine, philippines, banking, growth, digital, inclusion, financial inclusion, economic, payments, microfinance, market, inflation, banks, sentral, bangko, bangko sentral, strong, sector, support, external, philippine economy, reforms, ng, policy, sentral ng pilipinas, sentral ng, bangko sentral ng pilipinas, bangko sentral ng

Continued on next page

Topic	Label	Top 30 keywords
5	Fin. Reg.	financial, risk, capital, regulatory, banks, management, banking, regulation, supervision, risk management, risks, systemic, market, institutions, resolution, liquidity, requirements, crisis, basel, firms, standards, supervisory, large, stability, bank, need, credit, federal, financial institutions, federal reserve
6	CA	canada, inflation, canadian, bank canada, bank, economic, growth, economy, policy, global, target, expected, rate, monetary, demand, financial, monetary policy, canadian economy, prices, rates, household, trade, risks, inflation target, low, economic growth, investment, market, capacity, higher
7	Global	global, financial, capital, economies, emerging, international, countries, economic, trade, markets, flows, imf, exchange, policy, globalization, policies, capital flows, monetary, crisis, growth, emerging markets, stability, emes, domestic, imbalances, market, rates, global financial, current, exchange rate
8	AU	australia, growth, australian, rba, economic, inflation, rate, economy, rates, global, household, policy, investment, housing, low, high, expected, demand, market, debt, financial, increased, monetary, strong, prices, mining, cash rate, monetary policy, reserve bank, significant
9	SE	riksbank, inflation, policy, sweden, monetary policy, economic, monetary, rate, swedish, repo rate, repo, growth, target, stability, financial, low, prices, inflation target, market, price, wage, high, economy, rates, forecasts, household, crisis, term, potential, productivity
10	CH	snb, swiss, switzerland, franc, swiss franc, monetary, rate, policy, exchange, market, economic, monetary policy, national bank, banks, swiss national bank, swiss national, stability, foreign, rates, bank snb, national bank snb, swiss national bank snb, financial, inflation, national, price, economy, swiss economy, term, bank
11	ZA	south, africa, south africa, sarb, inflation, african, growth, economic, south african, global, policy, financial, monetary, exchange, monetary policy, rand, rate, bank, markets, reserve bank, stability, african reserve, south african reserve, south african reserve bank, african reserve bank, economic growth, domestic, market, foreign, reserve
12	Asia	asia, singapore, financial, asian, regional, asean, growth, mas, global, economic, trade, markets, market, development, bond, infrastructure, integration, capital, investment, china, region, sector, economies, cooperation, significant, like, financial sector, management, strong, opportunities
13	GB	uk, inflation, mpc, policy, economic, monetary, bank, growth, monetary policy, bank england, england, economy, demand, target, prices, rate, market, stability, financial, uk economy, global, rates, recovery, supply, policy committee, inflation target, significant, monetary policy committee, productivity, term
14	NO	norway, inflation, Norges, Norges bank, policy, rate, economic, petroleum, norwegian, high, low, fund, krone, growth, oil, monetary policy, monetary, rates, bank, prices, economy, wage, stable, government, revenues, employment, global, norwegian economy, exchange, output
15	ECB	euro, growth, euro area, area, ecb, price, economic, price stability, term, stability, rates, medium, risks, prices, inflation, expected, medium term, fiscal, monetary, structural, risks price, reforms, annual, structural reforms, hicp, risks price stability, hicp inflation, oil, council, governing council
16	Banking Union	banking, integration, european, financial, banks, supervisory, financial integration, union, single, supervision, ssm, banking union, market, euro, cross, capital, ecb, area, cross border, border, risk, euro area, markets, resolution, national, regulatory, eu, stability, mechanism, risks
17	IN	india, rbi, growth, market, financial, policy, sector, inflation, monetary, stability, indian, reforms, global, government, development, management, economic, markets, rate, foreign, exchange, fiscal, capital, investment, monetary policy, bank, significant, liquidity, infrastructure, external
18	Climate	climate, climate change, transition, green, risks, change, related, climate related, carbon, financial, finance, climate risks, net zero, banks, emissions, related risks, environmental, risk, sustainable, action, green finance, zero, disclosures, net, climate related risks, climate risk, nature, related financial, energy, climate related financial

Continued on next page

Topic	Label	Top 30 keywords
19	Islamic Finance	islamic, islamic finance, finance, islamic financial, malaysia, shariah, financial, sukuk, international, development, global, industry, islamic banking, institutions, market, growth, services, islamic financial institutions, islamic financial services, ifsb, banking, management, principles, conventional, financial services, products, innovation, growth islamic, sukuk market, financial institutions
20	NZ	zealand, new zealand, new, rbnz, reserve bank, bank, reserve, policy, financial, economic, inflation, housing, monetary, māori, monetary policy, stability, rate, bank new zealand, rates, reserve bank new zealand, exchange, growth, reserve bank new, bank new, ocr, high, prices, global, prudential, price
21	Fin. Stability	financial, stability, policy, financial stability, central, monetary, central banks, monetary policy, macroprudential, banks, crisis, price, prudential, macro, tools, policies, asset, macroprudential policy, price stability, macro prudential, bank, systemic, economic, risks, financial crisis, risk, need, crises, bubbles, inflation
22	HK	kong, hong kong, hong, rmb, mainland, financial, china, hkma, market, international, mainland china, bond, connect, offshore, dollar, international financial, exchange, markets, banking, development, hong kong financial, kong financial, hong kong dollar, kong dollar, offshore rmb, management, banks, currency, infrastructure, investors
23	Governance	risk, governance, management, risk management, compliance, corporate governance, corporate, internal, accounting, financial, directors, audit, practices, effective, standards, controls, risks, banks, board, auditors, value, internal controls, need, fair value, importance, business, fair, institutions, role, culture
24	TH	thailand, bank thailand, thai, financial, economic, growth, sector, policy, stability, thai economy, global, economy, bank, recovery, capital, term, investment, domestic, long term, monetary, long, private, bot, strong, master, master plan, infrastructure, external, banking, financial sector
25	Inclusion	inclusion, financial inclusion, financial, literacy, financial literacy, services, financial education, education, access, banking, rbi, india, credit, banks, financial services, products, accounts, technology, rural, poor, initiatives, bank, inclusive, groups, efforts, savings, financial products, crucial, protection, microfinance
26	Inflation	inflation, policy, expectations, monetary, monetary policy, economic, price, models, uncertainty, central, inflation expectations, stability, forecasts, term, making, central banks, output, understanding, price stability, long, decision, forecasting, decisions, prices, shocks, decision making, based, bank, rates, model
27	Housing	mortgage, community, housing, cra, federal, federal reserve, communities, reserve, community development, foreclosure, lending, credit, development, foreclosures, market, minority, subprime, income, borrowers, practices, loan, financial, data, access, cdfis, home, affordable, regulations, small, mortgage market
28	Banking	banks, banking, sector, financial, indian, india, credit, sector banks, management, regulatory, rbi, assets, public sector banks, psbs, risk, nbfc, npas, public sector, need, banking sector, governance, stressed, indian banking, bank, capital, public, risk management, growth, customer, ibc
29	ECB	ecb, pandemic, pepp, euro area, area, euro, measures, support, inflation, conditions, economic, monetary, recovery, policy, growth, fiscal, purchase, purchases, financing conditions, purchase programme, programme, billion, financing, monetary policy, term, asset, expected, 2020, 2021, favorable
30	CL	chile, inflation, central bank, policy, central, growth, monetary, economic, bank, external, monetary policy, financial, mpr, rate, global, expected, chilean, central bank chile, bank chile, conditions, domestic, monetary policy rate, peso, gdp, economy, exchange, cbc, fiscal, policy rate, risks
31	IL	israel, bank israel, growth, government, economic, bank, israeli, economy, stability, education, financial, low, poverty, gdp, policy, market, term, long, long term, budget, competition, labor, rate, public, global, israel economy, rates, israeli economy, credit, tax
32	ECB	inflation, ecb, energy, policy, monetary, euro area, monetary policy, area, euro, target, rates, supply, medium term, term, high, 2023, medium, economic, ukraine, pressures, wage, price, prices, transmission, rate, growth, expected, demand, data, data dependent

Continued on next page

Topic	Label	Top 30 keywords
33	DK	denmark, danish, nationalbank, danmarks, danmarks nationalbank, fixed exchange rate, fixed exchange, rate, exchange, exchange rate, fixed, fixed exchange rate policy, mortgage, policy, exchange rate policy, rate policy, rates, economic, financial, banks, euro, market, economy, fiscal, low, growth, danish economy, high, stability, government
34	Comms.	communication, transparency, central, policy, bank, central bank, expectations, public, guidance, monetary policy, central banks, monetary, forward guidance, forward, market, bank communication, central bank communication, effective, banks, economic, accountability, ecb, clear, inflation, effectiveness, market expectations, forecasts, rate, credibility, effective communication
35	Diversity	diversity, women, economics, gender, federal reserve, reserve, federal, community, education, economic, diverse, workforce, research, mobility, development, income, importance, inclusive, students, employment, work, labor, disparities, career, pandemic, programs, economists, communities, workforce development, opportunities
36	TR	turkey, inflation, cbtr, central bank, policy, turkish, central, stability, monetary, prices, global, price, bank, economic, monetary policy, food, rates, price stability, financial, rate, exchange, term, lira, demand, fiscal, growth, turkish lira, bank turkey, domestic, central bank turkey
37	Bench- marks	code, libor, market, fx, participants, transition, market participants, benchmarks, fx market, reference, contracts, global code, benchmark, risk free, sofr, rates, rate, reference rates, alternative, principles, 2021, based, free, global, overnight, sonia, trading, risk, robust, transactions
38	MY	malaysia, financial, sector, economic, banking, talent, growth, development, financial sector, global, malaysian, services, economy, industry, market, strong, new, islamic, negara, bank negara, institutions, domestic, ringgit, skills, negara malaysia, bank negara malaysia, management, future, regional, bank
39	Basel II	basel, basel ii, ii, risk, capital, banks, implementation, management, risk management, pillar, requirements, capital requirements, ii implementation, basel ii implementation, operational risk, supervisors, approaches, accord, banking, operational, supervisory, adequacy, capital adequacy, practices, advanced, irb, internal, framework, approach, regulatory
40	Liquidity	liquidity, market, banks, central, financial, central banks, bank, markets, risk, assets, funding, market liquidity, clf, lcr, central bank, management, crisis, liquidity risk, stress, stability, crises, liquid, risks, hqla, collateral, term, financial stability, like, lolr, regulatory
41	Insurance	insurance, industry, takaful, insurance industry, risk, insurers, financial, advisers, asia, financial advisers, products, malaysia, singapore, life, life insurance, development, takaful industry, management, new, growth, protection, regulatory, significant, advisory, financial advisory, consumer, penetration, solutions, business, risk management
42	AML	aml, laundering, money laundering, cft, aml cft, money, terrorism, crimes, financing, terrorism financing, compliance, financial, enforcement, financial crimes, anti, malaysia, anti money, anti money laundering, combating, institutions, international, tf, money laundering terrorism financing, laundering terrorism financing, laundering terrorism, money laundering terrorism, crime, fatf, ml, law
43	CN	china, financial, shanghai, market, development, opening, rmb, exchange, pbc, rate, reform, economic, foreign, support, exchange rate, enterprises, policy, growth, reforms, management, international, sector, china financial, center, monetary, monetary policy, rmb exchange, institutions, rmb exchange rate, bank
44	Data	data, statistics, statistical, analysis, financial, quality, collection, data collection, policy, importance, reporting, making, information, need, gaps, statisticians, economic, surveys, granular, rbi, ecb, challenges, detailed, central, area, euro area, monetary, including, bank, research
45	Solvency II	pra, solvency ii, solvency, insurers, ii, insurance, regulatory, risk, firms, uk, regulation, regime, capital, prudential, risks, ensure, policyholders, standards, policyholder, new, insurance sector, financial, prudential regulation, ma, requirements, matching adjustment, models, management, safety soundness, annuity
46	KR	korea, bank korea, bank, economic, korean, bok, financial, policy, global, stability, korean economy, growth, monetary, market, central, international, economy, monetary policy, financial stability, inflation, challenges, staff, foreign, external, crisis, cooperation, including, organizational, potential, recovery

Continued on next page

Topic	Label	Top 30 keywords
47	ECB	ecb, crisis, financial, liquidity, measures, stability, banks, support, fiscal, central, policy, term, exit, non standard, credit, euro, monetary, standard, central banks, euro area, price stability, area, financial crisis, market, bank, price, non standard measures, standard measures, economic, operations
48	Community Banks	community, community banks, banks, regulatory, federal reserve, federal, reserve, banking, local, small, community banking, bank, lending, supervisory, community bankers, regulations, community bank, communities, supervision, bankers, compliance, risk, burdens, challenges, credit, new, size, larger, burden, role

Note. Keywords are ordered by probability (importance), with higher probability terms listed first. The labels are manually assigned based on the top keywords for interpretive clarity. "Generic" topic (-1) refers to speeches that could not be confidently assigned to any specific topic.

C Prompts used for analysis

C.1 Prompt used to summarise speeches

- **Objective:** You are tasked with summarizing the key points made in speeches given by central bankers, related to central banking. This aids in research and retrieval efficiency.
- **Task:** For each speech document provided, summarize the speech in bullet points, as many as are necessary to capture all the key messages and rationales contained in the speech.
- **Ordering:** Arrange the bullet points in order of their importance as reflected in the document.
- **Formatting:** Present your bulleted key points in a list, ordered by priority. Return only bullet points and nothing else in your answer.

C.2 Prompt used for global topic labelling

You are assisting with labeling clusters of central bank speeches based on keyword analysis. Each cluster is represented by a list of n-gram keywords (ranked by importance via c-TF-IDF). Your task is to assign a **concise, human-readable label** to each cluster that reflects the overall theme or source of the speeches. Please follow these guidelines:

- Output a **single label** for each keyword list (1–2 words maximum).
- Labels must be appropriate for **graph axes, cluster visualizations, or dashboards**.
- Emphasize the **top-ranked keywords** in determining the label.
- Labels can be:
 - A **country code** (ISO 2-letter, e.g., “US”, “GB”, “IN”)
 - A **region** (e.g., “Asia”, “Global”)
 - A **policy or thematic area** (e.g., “Climate”, “Liquidity”, “Payments”)
 - An **institution or policy framework** (e.g., “ECB”, “Basel II”, “Fin. Reg.”)
- Avoid explanatory text, quotes, or full sentences.
- Only return a table with a **single column**: the label for each corresponding input row.

C.3 Prompt used for ECB topic labelling

You are assisting with labeling clusters of European Central Bank speeches based on keyword analysis. Each cluster is represented by a list of n-gram keywords (ranked by importance via c-TF-IDF). Your task is to assign a **concise, human-readable label** to each cluster that reflects the overall theme or source of the speeches. Please follow these guidelines:

- Output a **single label** for each keyword list (2–3 words maximum).
- Labels must be appropriate for **graph axes, cluster visualizations, or dashboards**.
- Emphasize the **top-ranked keywords** in determining the label.
- Labels can be a **policy or thematic area** specific to the central bank (e.g., “Monetary Policy”, “Financial Stability”, “Supervision”, “Eurozone Integration”, “Liquidity”, “Payments”).
- Avoid explanatory text, quotes, or full sentences.
- Only return a table with a **single column**: the label for each corresponding input row.

C.4 Prompt used for monetary policy sentiment analysis

- **System prompt:** You are an Economist specializing in monetary policy. Your task is to assess and classify the monetary policy sentiment of summarized central banker speeches.
- **User prompt:** Please assess the hawkishness or dovishness of a central banker's speech based on monetary policy sentiment. Use a score of 1, 0, or -1, where: 1 indicates hawkishness (suggesting tightening measures like rate hikes or balance sheet reductions), 0 is neutral, and -1 indicates dovishness (suggesting loosening or accommodative policies). Provide specific language or key points from the speech that influenced your assessment, in bullet points. Keep your response concise.

Previous volumes in this series

1298 October 2025	Exploring household adoption and usage of generative AI: new evidence from Italy	Leonardo Gambacorta, Tullio Jappelli and Tommaso Oliviero
1297 October 2025	The BIS multisector model: a multi-country environment for macroeconomic analysis	Matthias Burgert, Giulio Cornelli, Burcu Erik, Benoit Mojon, Daniel Rees and Matthias Rottner
1296 October 2025	Predicting the payment preference for CBDC: a discrete choice experiment	Syngjoo Choi, Bongseop Kim, Young Sik Kim, Ohik Kwon and Soeun Park
1295 October 2025	Pricing in fast payments: a practical and theoretical overview	Jose Aurazo, Holti Banka, Guillermo Galicia, Nilima Ramteke, Vatsala Shreeti and Kiyotaka Tanaka
1294 October 2025	Parsing the pulse: decomposing macroeconomic sentiment with LLMs	Byeungchun Kwon, Taejin Park, Phurichai Rungcharoenkitkul and Frank Smets
1293 October 2025	International risk sharing and wealth allocation with higher order cumulants	Giancarlo Corsetti, Anna Lipińska and Giovanni Lombardo
1292 September 2025	Macroeconomic impact of weather disasters: a global and sectoral analysis	Torsten Ehlers, Jon Frost, Carlos Madeira and Ilhyock Shim
1291 September 2025	Harnessing artificial intelligence for monitoring financial markets	Matteo Aquilina, Douglas Araujo, Gaston Gelos, Taejin Park and Fernando Pérez-Cruz
1290 September 2025	The speed premium: high-frequency trading and the cost of capital	Matteo Aquilina, Gbenga Ibikunle, Khaladdin Rzayev and Xuesi Wang
1289 September 2025	The liquidity state dependence of monetary policy transmission	Oliver Ashtari-Tafti, Rodrigo Guimaraes, Gabor Pinter and Jean-Charles Wijnandts
1288 September 2025	The capital puzzle	Eduardo Amaral
1287 September 2025	Banks' regulatory risk tolerance	Mikael Juselius, Aurea Ponte Marques and Nikola Tarashev
1286 September 2025	How do quantitative easing and tightening affect firms?	Egemen Eren, Denis Gorea and Daojing Zhai

All volumes are available on our website www.bis.org.