EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Ami Dalloul, Anisa Tiza Mimun

ECB perception of non-banks: a machine learning-based approach

No 3304 / November 2026

Disclaimer: This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

Abstract

This paper studies the European Central Bank's (ECB) perception of non-bank financial intermediaries (NBFIs) and how this perception is publicly communicated. We construct a perception index based on ECB speeches over the period 1998–2025. Sentiment in references to non-banks is classified by a large language model that we fine-tune on our own hand-labeled dataset constructed for this purpose. The resulting index reveals a predominantly negative and declining perception over time, particularly since the Global Financial Crisis. Negative sentiment is primarily associated with concerns related to financial stability and regulatory needs, whereas positive sentiment reflects the role of non-banks in fostering financial innovation and supporting market functioning. The index is made publicly available for future research.

Keywords: ECB speeches, non-bank financial intermediation, machine learning, sentiment analysis, large language models.

JEL classification: E44, E52, C80

Non-technical summary

This paper studies how the European Central Bank (ECB) perceives non-bank financial intermediaries (NBFIs), often referred to as non-banks, and how this perception is communicated to the public. The non-bank sector, which includes entities such as investment funds, insurance companies, and other financial intermediaries outside the traditional banking system, has grown significantly over the past decade, and with it, growing attention from researchers and policy-makers to its role in the economy, including its involvement in monetary policy transmission and its potential to pose financial stability risks. Despite this growing body of work, relatively little is known about how institutions with policy-making mandates actually perceive non-banks. This paper aims to fill that gap by introducing a way to quantify the ECB's perception and by documenting how that perception has evolved.

Against this backdrop, we construct a perception index based on ECB speeches delivered between 1998 and 2025. We identify all sentences in these speeches that refer to the non-bank sector and classify the sentiment expressed in each reference, i.e. whether positive, negative, or neutral, using a large language model. Because no existing tool is designed to perform this task accurately for our specific context, we fine-tune the model ourselves, training it on a dataset of 1,300 sentences that we manually labeled for this purpose. The resulting model reaches a good level of accuracy, allowing us to reliably track how the ECB's tone toward non-banks has changed over nearly three decades.

Our resulting index shows that the ECB's perception of non-banks has been predominantly negative and has become more so over time, particularly since the Global Financial Crisis (GFC). While the overall trend declines, we do observe episodic periods of more positive sentiment, most notably in the years before the GFC and again around 2014-2015, when European authorities were implementing regulatory reforms for the non-bank sector following the crisis.

Beyond simply measuring the tone of ECB communication, we examine the underlying reasons behind these shifts in perception. We find that negative sentiment is closely tied to concerns about financial stability and the need for stronger regulation, for instance, worries about liquidity, leverage, and the ways in which

non-banks are interconnected with traditional banks, which could allow financial shocks to spread across the system. More recently, negative sentiment has also been associated with concerns about crypto-assets and decentralized finance, as well as gaps in the data available to monitor the sector. Positive sentiment, by contrast, tends to emerge when the ECB highlights the contributions of non-banks to financial innovation, market integration, diversification, and the smooth functioning of financial markets more broadly.

This paper makes two main contributions available to the broader research community. First, we introduce the first index, to the best of our knowledge, that quantifies how a major central bank publicly perceives the non-bank financial sector, offering a new tool for studying central bank communication. Second, we make the resulting index available to other researchers, upon request, for use in future empirical work on central bank communication, non-bank finance, and monetary policy transmission.

1 Introduction

The non-bank financial intermediaries (NBFI) sector, often referred to as non-banks, has grown significantly in the last decade. As total assets have surged, the relevance of this heterogeneous and complex sector has increased in parallel. Researchers and policy-makers have devoted increasing attention to understanding its main dynamics, its role in the transmission of monetary policy, and the potential financial stability risks it may pose (Giuzio et al., 2021, Holm-Hadulla et al., 2023, Jin and Nadal De Simone, 2020, Kaufmann, 2023, Tiza Mimun et al., 2025). Despite this growing body of work, relatively little is known about how institutions, in particular those with policy-making mandates, actually *perceive* non-banks. Our constructed index aims to fill this gap in introducing how the ECB perception actually is and, most importantly, how the perception can actually be quantified. We construct a machine learning-based index which reflects the perception of the ECB about non-banks, aggregated at the monthly level. To the best of our knowledge, no comparable measure exists in the literature. Our contribution is threefold. First, we introduce a perception index that quantifies the ECB's publicly expressed views of non-banks. The index is constructed from publicly available ECB speeches covering the period 1998 to 2025. To extract sentiment, we classify all sentences referring to the NBFI sector using a pretrained large language model, which we fine-tuned with our manually labeled dataset. Overall, our constructed index leans negatively with some positive spikes. However, we observe a gradual downward trend over time. At the yearly level, the index becomes almost exclusively negative from the great financial crisis onward. Looking on a quarterly level, some episodic improvements in perception can be seen, but these do not offset the broader tendency toward more negative perceptions.

Second, we explore the narratives behind the ECB's perception of non-banks. The speeches address a wide range of topics, including regulation, market integration, growth, diversification, securitization, risk, financial stability, innovation, credit, and monetary policy. Negative perception arises largely when the ECB discusses the need for tighter regulation, concerns about liquidity and leverage, or the possibility that interconnectedness with banks may create channels for contagion, increased risk taking or competition pressures. Positive sentiment is more common

when the ECB highlights the contribution of non-banks to financial innovation, market functioning, and risk-bearing capacity in the real economy. Third, we make both the labeled data and the resulting index available to other researchers. The manually labeled dataset can be used to fine-tune alternative classifiers, or to serve as a ground truth for benchmarking the accuracy of other models on central bank speeches in the context of NBFIs. The index itself can be used as an input in empirical work on central bank communication, non-bank finance, and monetary policy transmission. Both are available upon request.

Robustness checks show that the model selected for constructing the index consistently outperforms alternative approaches across a wide range of fine-tuning specifications, including variation in context size, reference identification, and class-imbalance treatments. The index is also stable when computed using two different aggregation methods, in line with the literature. Moreover, the three best-performing models, a fine-tuned RoBERTa classifier, and two fine-tuned Llama-3 (one operating at the sentence level and one using a context window), exhibit very similar dynamics. Their close co-movement indicates that the index captures a coherent perception.

The remainder of the paper is structured as follows. Section 2 reviews the related literature, focusing on machine-learning-based indices and central bank communication. Section 3 describes the text dataset, the fine-tuning process, and the construction of the perception index. Section 4 describes the narratives behind the perception. Section 5 presents the robustness checks. Section 6 concludes.

2 Related Literature

Following the first and second wave of increased central bank (CB) communication (Haldane et al., 2020), a large body of literature on its effects has emerged. A significant portion of this research focuses on two primary target audiences: market participants and the general public comprising households and firms. Studies concerning households and firms have found that although these agents stand to benefit from receiving central bank information (Dräger and Nghiem, 2023, Nghiem et al., 2024), empirical evidence suggests that their responsiveness is limited (Coibion et al., 2018), despite functioning media transmission channels (Mun-

day and Brookes, 2021).

In contrast, market participants demonstrate a higher degree of attentiveness and reactivity in response to central bank communications. High-frequency financial markets and asset prices have been proven to react to news coming from central banks about monetary policy decisions, forecasts and forward guidance (Cai et al., 2021, Born et al., 2014, Ehrmann and Talmi, 2020, Gómez-Cram and Grotteria, 2022, Gürkaynak et al., 2005, Mumtaz et al., 2023, Schmeling and Wagner, 2024). More recently, Altavilla et al. (2025) document that inter-meeting communication by ECB Executive Board members, including speeches and interviews outside of Governing Council meetings, generates significant intraday market movements, often of similar or larger magnitude than formal policy announcements. This finding underscores the growing importance of speech-based communication as a policy instrument in its own right.

A substantial portion of the literature on central bank communications has increasingly relied on computational linguistic methods. Early contributions employed dictionary-based approaches and manual coding to quantify the tone and sentiment of policy statements (Ehrmann and Fratzscher, 2007, Gürkaynak et al., 2005). Building on these initial efforts, more recent studies have integrated advanced natural language processing techniques, such as topic modeling, machine learning algorithms for sentiment analysis, and vector space models, to extract subtler nuances from central bank texts (Aruoba and Drechsel, 2022, Coibion et al., 2018, Gürkaynak et al., 2005, Haldane et al., 2020). Parle (2022) constructs hawk-dove indices from ECB press conference transcripts using text-based methods and documents their effects on financial markets, while Kaminskas and Jurkaias (2024) build sentiment indices from ECB communication events and examine their relationship with the broader economic environment and asset prices. A broader survey of text mining methods applied to central bank communication is provided by Bricongne et al. (2024), who review the expanding toolkit available to researchers and policymakers for quantifying the informational content of central bank language. These methodological advances have provided deeper insights into how central bank communications shape expectations and behavior across diverse audiences, including market participants, households, and firms.

3 Text Analysis Methodology

In this section, we describe the steps undertaken to construct our index. We explain the choice of LLMs as a method for quantifying text. We present the data labeling process as well as the fine-tuning and validation process. We finally explain the calculation of the perception time series.

To build our ECB NBFI perception index, we collect 2,830 speeches given by the ECB between July 1998 and June 2024. The sentences within these speeches are then first classified according to their relevance to the NBFI sector, and subsequently, each sentence is assigned a corresponding perception/tone score.

3.1 Quantifying Text

A popular approach to labeling text is unsupervised machine learning methods such as the Latent Dirichlet Analysis (LDA) method (Blei et al., 2003). Topic models capture the various topics that can occur within a set of documents (corpus). This is achieved by observing the co-occurrence between words in a sentence (or paragraph) and clustering them. However, such models neither allow for searching for pre-specified topics, nor do they output an interpretable sentiment for the uncovered topics.

Measurement of the tone of text to estimate the effects of communication on various economic agents has become increasingly prevalent. A highly used method of sentiment analysis has been dictionary methods (Apel and Blix Grimaldi, 2012, Bennani, 2020, Bennani and Neuenkirch, 2017, Correa et al., 2021, Loughran, 2011, Picault and Renault, 2017, Schmeling and Wagner, 2024, Stekler and Symington, 2016). When using a dictionary method, a set of n-grams (words) is predefined, and a document's sentiment score is calculated based on the relative occurrence of positive (e.g. 'good') and negative (e.g. 'bad') n-grams. However, as Pfeifer and Marohl (2023) point out, the bag-of-word approach struggles to address the linguistic idiosyncrasies of central bank communications. Indeed, central bank communication tends to depend on context, which n-grams cannot properly capture. For instance, the bigram "lower interest" will be counted as occurring in both the sentences "we will not lower interest" and "we will lower interest" even though

they mean the opposite of each other.
In the case of sentiment analysis, deep-learning models (also referred to as Large Language Models (LLM)) have been found to outperform both dictionary methods (Fischer et al., 2023, Frankel et al., 2022, Purda and Skillicorn, 2015) and non-deep-learning machine learning models (Huang et al., 2022).
LLMs are understood to be typically pre-trained on large text datasets (Peters et al., 2018, Radford, 2018). They offer the advantage of considering the context of a word within a specific sentence, allowing them to bypass the challenges faced by dictionary methods. Furthermore, LLMs can be fine-tuned to accommodate a specific dataset, making it possible to predict sentiment for out-of-sample data. Fischer et al. (2023) use a human annotated dataset to fine-tune a FinBert model on FOMC meeting transcripts to label sentences with a set of pre-defined topics and to assign them a corresponding sentiment.

3.2 Training data and labeling methodology

Our corpus is made up of 2,830 ECB speeches over a time span of nearly three decades from 1997 until 2025. We follow Pfeifer and Marohl (2023) and keep preprocessing to a minimum as we are using LLMs that do not require the removal of stop words, punctuation, stemming, or lemmatizing. We thus split our corpus into sentences that will be used as input for the model.
After sentence tokenization, our full text data frame contains around 315,400 sentences. To construct the perception index, we first identify sentences as either referring to the NBFI sector or addressing other non-NBFI-related topics. Subsequently, each sentence that refers directly or indirectly to non-banks is assigned a label that indicates whether it communicates a positive, negative, or neutral perception. Given the large number of sentences, manual labeling is time prohibitive. We thus randomly select 4,000 sentences from the full dataset to be manually labeled, used as a benchmark, and for training in order to choose the best performing model for topic and tone identification.
Figure 1 displays the words found in the sentences that capture the ECB's perception of the NBFI sector. Meanwhile, Figure 2 illustrates the specific words that occur most frequently in sentences assigned negative, neutral, or positive per-

ception labels. When examining all text related to the NBFI sector, we observe expected terms such as non-bank, market, assets, credit, financial, risk, funds, and policy. Similarly, when analyzing the words that correlate with negative versus positive perception labels, Figure 2 reveals that negative perception sentences are often associated with terms such as risky, amplify, pressure and illiquid, while positive perception sentences tend to feature associations with words such as encourage, foster, and positive. These aggregate frequencies, however, mask considerable overlap. Terms such as risk, leverage, and regulation appear in sentences across all three labels, and their presence alone does not determine the perception conveyed.



Figure 1: Wordcloud of Non-Bank Related Sentences in ECB Speeches

Note: This figure presents a word cloud generated from sentences labeled as belonging to the topic of the NBFI sector in speeches by the ECB between 1998 and 2024. The word clouds were derived using TF-IDF vectorization, which identifies key terms within each sentiment group based on their frequency and relevance.

When constructing our training dataset, we assign two labels to each sentence. A topic label classifying the sentences as either referring to the NBFI sector or as "other" for all other topics, as well as a perception/tone label for those sentences that refer to non-banks. Sentences can be labeled as positive (1), neutral (0), or negative (-1).

As a broad definition, a neutral tone is a statement about the NBFI sector that states facts without obvious negative or positive descriptive adjectives. In order to capture opposite perceptions depending on different perspectives, i.e. negative or positive, we draw on different conclusions and their respective interpretation in the literature. For example, a board member may speak about non-banks in the

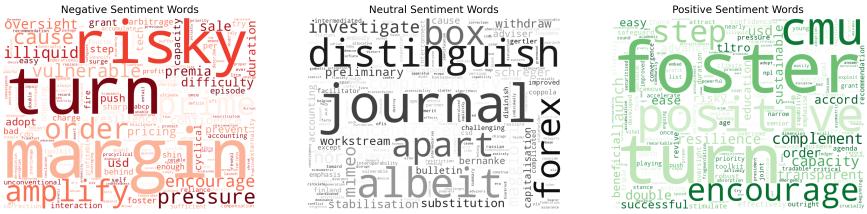


Figure 2: Wordclouds by Sentiment of the Perception of the ECB of the NBFI Sector

Note: This figure presents word clouds generated from text data related to the perception of non-bank financial intermediaries (NBFIs) across different tone of perception categories: Positive, Neutral, and Negative. The word clouds were derived using TF-IDF vectorization, which identifies key terms within each sentiment group based on their frequency and relevance.

context of rising bank-non-bank interconnections and resulting financial stability concerns (see for example Franceschi et al. (2023)). We argue that this would most likely reflect a negative perception of the ECB regarding this specific aspect, and therefore, it would reflect an overall negative perception. Further, a negative perception could also reflect, for example, concerns regarding overall challenges for the transmission of monetary policy through non-banks. For this assumption, we follow the literature on the transmission of monetary policy through banks and non-banks. This literature examines whether and how non-bank lending increasingly substitutes for bank lending in response to monetary policy (Anyfantaki et al., 2026, Hodula and Tiza Mimun, 2025, Bednarek et al., 2025, Tiza Mimun et al., 2025). Conversely, ECB board members may communicate a positive perception of non-banks when they discuss, for example, the role of non-banks in improving firms' and households' access to financing, or in helping them diversify their funding sources.

Defining perception in this way also shows why a dictionary approach is not suitable here. The problem is not only the lack of context discussed in Section 3.1, but that the same term can signal opposite perceptions depending on how it is used. A stated need for regulation of the NBFI sector points to perceived vulnerabilities, whereas a reference to regulation already in place points to a safeguard

against them. A sentence describing a shift from higher to lower risk assets conveys a positive perception, while reversing the direction of that shift would leave the same set of words and convey the opposite. A predefined word list cannot separate these cases, whereas manual labeling combined with fine-tuning allows the model to learn the distinction from how the terms are used in ECB speeches.
To translate these definitions into labeling decisions, we follow a set of rules for tone classification that is specific to what we observed in the data during the labeling process.

1. Sentences that contain references to the transmission of monetary policy through non-banks and explicitly raise concerns about its functioning are labeled as negative.

2. Sentences that voice concerns about interlinkage between banks and non-banks in relation to contagion risks are labeled as negative.

3. Sentences that contain references to a specific need for regulation of the NBFI sector are labeled as negative, as they could be a sign of increasing financial instability.

4. Sentences that discuss lack of data on NBFI and the subsequent consequences in terms of risk analysis, stability, monetary transmission appreciation, estimation and evaluation of size and reach of the sector, receive a negative label.

5. Sentences that contain mentions of already implemented regulation of the NBFIs are labeled as positive as they might reflect a guaranteeing of financial stability.

6. Sentences that discuss the positive impact of NBFIs in credit supply, financial innovation and market integration are assigned a positive label.

7. Sentences that describe any growth in the size of the NBFI sector without any explicit descriptive adjectives are labeled as neutral.

8. Sentences that describe the functioning of NBFIs or general market conditions in relation to NBFIs without explicit tone adjectives are labeled as neutral.

Table 1 shows example sentences and their corresponding labels.

Sentence	**Perception**	**Topic**
At the same time, the increase in non-bank finance has created new hazards for monetary policy.	-1	Non Bank
The investment fund industry doubled in size in the last five years in Europe.	0	Non Bank
Fund investors shifted their exposure from higher to lower risk assets, especially from high-yield corporate to government bond funds.	1	Non Bank
Persistent ultra-low policy interest rates may not be the right medicine for lifting inflation close to the target.	N.A.	Other

Table 1: Sample of manually labeled sentences from the training dataset

Of the 4,000 labeled sentences, we found that 1,300 referred to the NBFI sector, with negative sentiment being more prevalent than the other two sentiments. These 1,300 sentences form the dataset used to fine-tune the models for perception labeling. We then split them into training, testing, and validation subsets. The training set is used to train the different models. The validation subset is used to check the improvement of the model after each training round. Finally, the testing subset is used to check the final performance of the model post fine-tuning. By conducting this split, we ensure no data leakage, such that the model is not tested on data it has seen during training.

For the topic labeling task, we found that GPT-4o-mini outperformed both BERT and Llama2. Its accuracy rate, which measures the proportion of total predictions that are correct, was at 0.93. Since it required no fine-tuning, we ran the entire corpus through GPT-4o-mini, and labeled all sentences that related to the NBFI sector. Of the roughly 300,000 sentences in our corpus, 15,017 were labeled as "non_bank".

For the tone of perception labeling task, we looked at various LLMs without any

fine-tuning to establish a baseline for their performance off-the-shelf. We ran FinBERT, RoBERTa, Llama2 in version 7b-hf and 13b-hf as well as Llama3-8B. All of these performed poorly in the off-the-shelf versions. Models such as FinBERT are trained on financial news and analyst reports, where sentiment reflects favorable or unfavorable news for market participants rather than the perception of a sector by a policy institution. Our labeling scheme rests on domain-specific distinctions that a model cannot infer without being trained on examples.

3.3 Fine-tuning for sentiment

To automate the labeling of the 15,017 non-bank sentences, we fine-tuned 5 models: two classifiers, FinBERT and RoBERTa, and three open source LLMs, Llama2 7b-hf, Llama2 13b-hf and Llama3-8B. Fine-tuning involves updating the pre-trained model's parameters on our manually labeled dataset, allowing the model to adapt its representations to the domain-specific language of ECB speeches and the task of perception classification. The Llama2 7b-hf contains 7 billion parameters, whereas Llama2 13b-hf includes 13 billion parameters, offering a larger capacity for processing and generating more nuanced text. Llama3 is, at the time of writing, one of the newer models from Meta.

For the large language models, we fine-tune two variants for sentiment classification: one operating at the sentence level and another using a context window. In the window-based setup, the model receives the target sentence together with its preceding and following sentences to provide contextual information. In the sentence-level setup, the model sees only the sentence to be classified. FinBERT and RoBERTa are trained in the same sentence-level manner, as they do not support context-window inputs.

In their baseline form, the different models perform rather poorly. In Table 4 we report the results of the top performing models. The highest weighted average precision is 0.52 for the off-the-shelf Llama3-8B model in both sentence and window context. The highest weighted F1 score is 0.26 for Llama3-8B at the sentence level. Fine-tuning improves performance for most models. The top two performers shown in Table 4 have an F1 score ranging from 0.70 to 0.80 after fine-tuning. The details of all other specifications such as hyperparameters and confusion matrices for each

of these three models can be found in Appendix A. The details of all other models we tested are further discussed in the robustness checks section.
For the calculation of our main perception index, we chose the Llama3-8B with a context window of 3 sentences as it showed the best post-fine-tuning metrics.

Model	Precision	Recall	F1
Llama3-8B Baseline (sentence)	0.52	0.32	0.26
Llama3-8B fine-tuned (sentence)	0.69	0.70	0.70
Llama3-8B Baseline (window-3)	0.52	0.26	0.19
Llama3-8B fine-tuned (window-3)	0.79	0.80	0.80

Table 2: Results of top two models before and after fine-tuning

Note: Precision, recall, and F1 are weighted averages across the three perception classes, computed on the held-out test set.

3.4 Sentiment Time Series

We follow Correa et al. (2021) by aggregating the sentence-level perception score in order to obtain a speech-level perception score.

$$\text{Perception}_{d,t} = \frac{\sum\{\text{positive sentences}_{d,t}\} - \sum\{\text{negative sentences}_{d,t}\}}{\sum\{\text{total sentences}_{d,t}\}} \quad (1)$$

Equation 1 shows how we extract the aggregated perception from the sentence-level labels predicted by the model. For each time period t (day, quarter, or year) in which speech(es) d are delivered, the index is calculated as the ratio of the total number of positive sentences minus the total number of negative sentences to the total number of sentences contained in the document(s).
Figure 3 shows the sentiment time series plotted over years. In the years prior to the great financial crisis (GFC), the perception index shows a few years where positive perception of the NBFI can be measured. A potential explanation, further mirrored in section 4, is that the creation of the monetary union was accompanied with enthusiasm about increased market access, growth and integration with

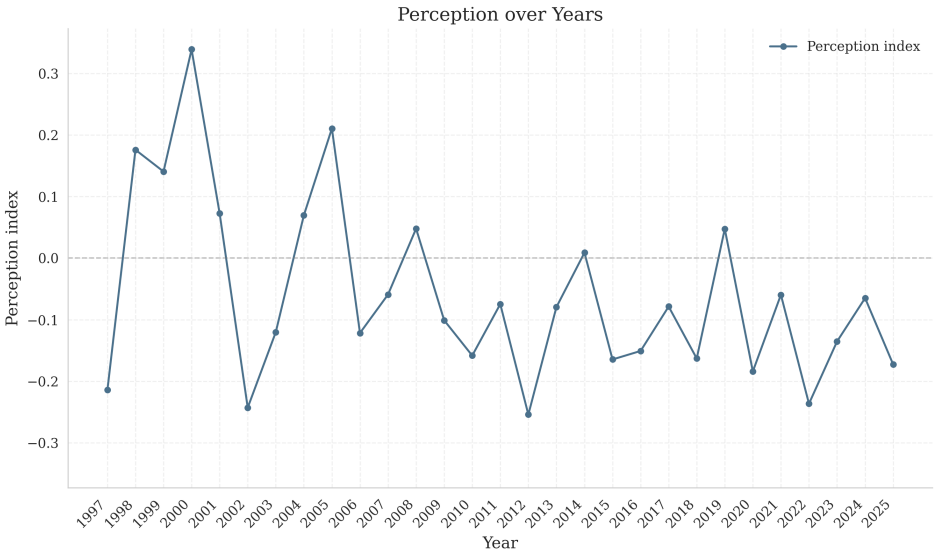


Figure 3: ECB-NBFI Perception Index — Yearly

Note: In this figure we plot the time series for the perception index, aggregated over years.

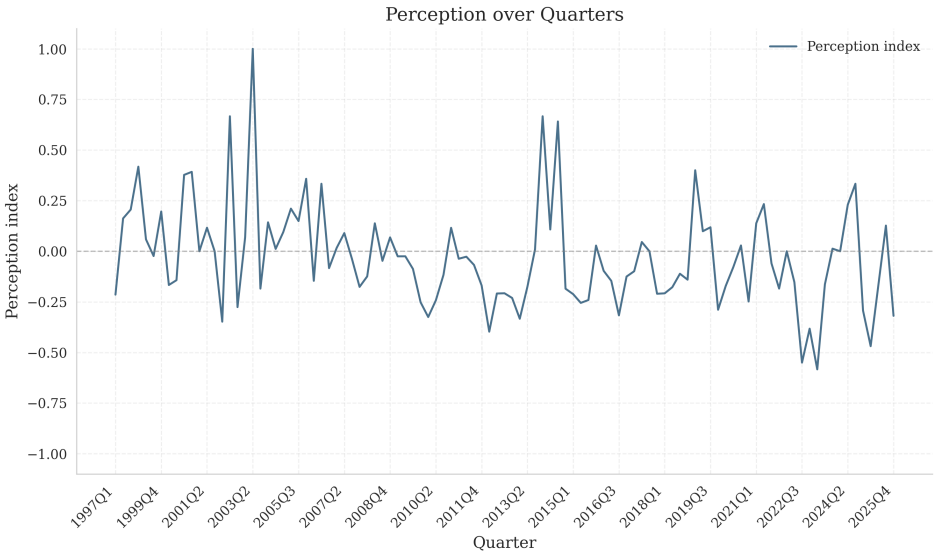


Figure 4: ECB-NBFI Perception Index — Quarterly

Note: In this figure we plot the time series for the perception index, aggregated over quarters.

non-bank entities having the potential to positively contribute to this. However, the index shows an overall negative trend with low levels at -0.25 such as during the European sovereign debt crisis. With the implementation of post-GFC regulatory frameworks, the perception index rises again, reaching more positive values after 2012. Despite this recovery, the index fluctuates significantly over the following years, with another period of negative perception emerging around 2016–2020, which corresponds to concerns over systemic risks and vulnerabilities within the NBFI sector. Figure 4 shows the sentiment time series plotted over quarters. By plotting it over shorter time intervals, we are able to see more of the within-year volatility of the perception. The positive spikes observed around the second quarter of 2013 mirror the enthusiasm with implementation of post-GFC regulatory frameworks. Despite this recovery, the index fluctuates significantly over the following years, with further periods of negative perception which correspond to concerns over systemic risks and vulnerabilities within the NBFI sector.

4 Narratives behind the Perception

Figure 5 offers further insight into the perception index by presenting a timeline that groups various topics found in the ECB speeches according to their associated perception and the corresponding time period. We cluster sentences by periods where we observe positive perception spikes or negative perception spikes. We then extract the most common topics found within these separate positive and negative clusters. This allows us to gain insight into what is behind the sign of the perception, and into the economic mechanisms our labeling rules are designed to capture (Rules 1-8, Section 3.2). [1]

In the earliest period covered by our corpus (1998-2001), the dominant topics are market integration, market access and growth, and diversification and competition. These map closely onto Rule 6, under which we label sentences discussing the contribution of non-banks to credit supply, financial innovation, and market in-

[1] We refer to the numbered classification rules introduced in our labeling methodology: concerns about monetary policy transmission (Rule 1), bank-non-bank interconnectedness and contagion (Rule 2), calls for tighter regulation (Rule 3), data gaps (Rule 4), already-implemented regulation (Rule 5), positive contributions to credit supply, innovation and market integration (Rule 6), and neutral descriptions of growth or market functioning (Rules 7-8).

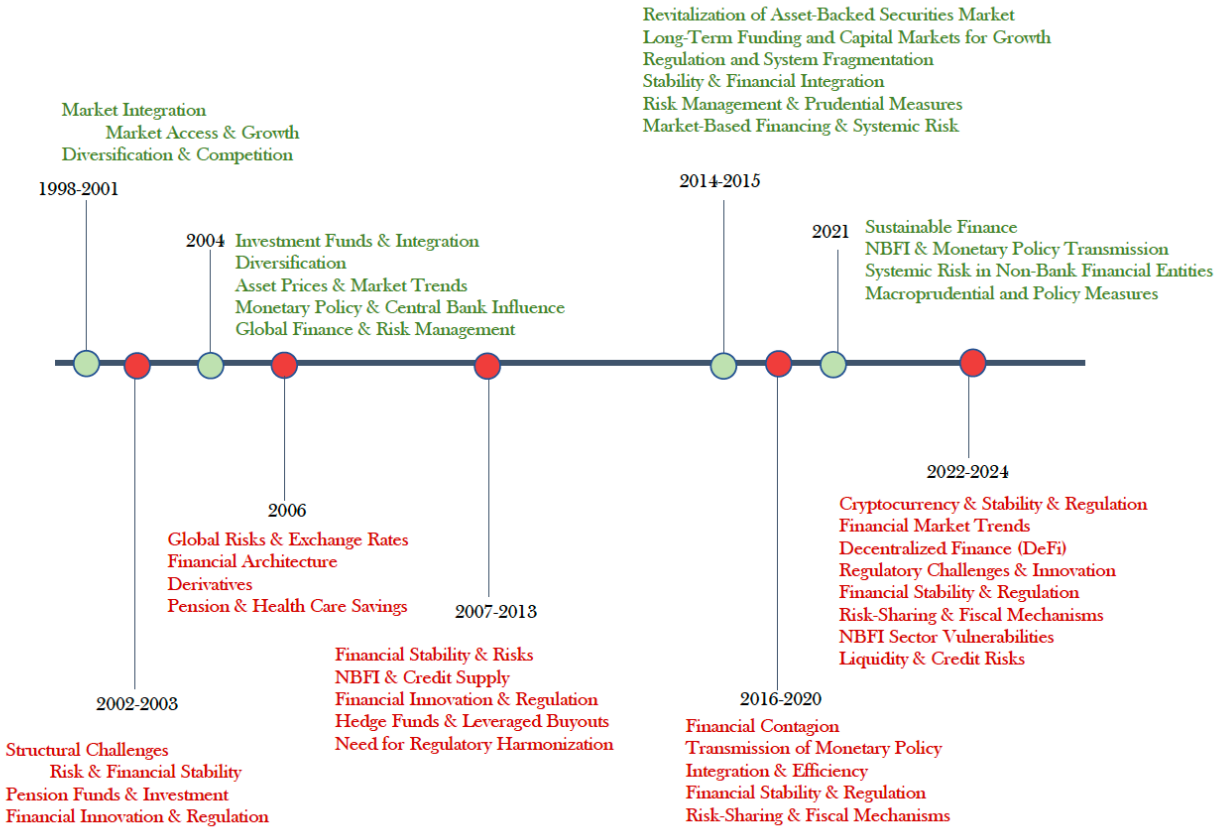


Figure 5: Clusters of Positive and Negative Topics Discussed in ECB-NBFI Perception Index over Time

Note: In this figure we show topics that can be found in the ECB speeches clustered by perception and time period. Topics relating to a negative perception are found below the line while positive ones are above it.

tegration as positive. In the run-up to and immediately following the introduction of the euro, ECB communication emphasized how non-bank intermediaries broadened firms' and households' access to financing beyond the bank lending channel, consistent with the view that a more diversified financial structure improves the allocation of capital and reduces the economy's dependence on any single type of intermediary. By 2002-2003, however, the tone shifts: topics move toward structural challenges, risk and financial stability, pension fund investment behavior, and financial innovation and regulation. This shift coincides with the aftermath of the dot-com crash and corporate governance concerns of the early 2000s, and is captured under Rule 3 (calls for tighter oversight) and Rule 2 (concerns about the propagation of shocks through interconnected balance sheets), as ECB communication increasingly framed the growing size and risk-taking of pension funds and other institutional investors as a potential source of instability rather than purely as a diversification benefit.

The 2004 cluster returns to positive territory, with topics centered on investment funds, integration, diversification, asset prices, the role of monetary policy and central bank influence, and global finance and risk management, again consistent with Rule 6 and reflecting a period in which non-bank credit intermediation was widely viewed as supporting market completeness and risk-sharing. The tone turns negative again in 2006, when topics shift toward global risks and exchange rates, financial architecture, derivatives, and pension and health-care savings. This period predates the Global Financial Crisis (GFC) but already reflects growing unease with the increasing complexity of the financial architecture and the expanding use of derivatives by non-bank intermediaries, an economic mechanism we label under Rule 2 and Rule 4: derivatives and increasingly complex balance-sheet structures raise concerns about interconnectedness and about the ability of supervisors to adequately monitor risk given limited data on non-bank activity, foreshadowing concerns that would become central after 2007 and consistent with the pre-crisis literature on the procyclicality of non-bank, market-based intermediaries' balance sheets, in which leverage and asset growth move together over the cycle and amplify swings in credit provision (Adrian and Shin, 2010).

The negative perception shift is most pronounced between 2007 and 2013, with topics dominated by financial stability and risks, NBFI and credit supply, finan-

cial innovation and regulation, hedge funds and leveraged buyouts, and the need for regulatory harmonization. These map directly onto Rules 1-4: sentences raising concerns about the transmission of monetary policy through non-banks (Rule 1), interconnectedness and contagion risk between banks and non-banks (Rule 2), explicit calls for regulation (Rule 3), and data gaps hindering risk assessment (Rule 4) are all classified as negative, and it is this cluster of concerns that dominates ECB communication during and after the crisis. Liquidity provision, as documented in the literature, depends on the availability of funding, and when funding becomes constrained, forced deleveraging by market participants can generate self-reinforcing liquidity spirals in which market liquidity and funding liquidity mutually reinforce each other's deterioration (Brunnermeier and Pedersen, 2009); applied to non-bank intermediaries, whose short-term, market-based funding structures make them particularly sensitive to this dynamic, this mechanism helps explain why the growth of the sector in the pre-crisis years, driven by procyclical leverage that expands during booms and contracts sharply in downturns, left it particularly exposed once market liquidity deteriorated (Adrian and Shin, 2010). This period also marks the point at which international bodies began to coordinate a policy response, with the Financial Stability Board finalizing a policy framework for the oversight and regulation of non-bank entities in 2013, subsequently endorsed by the G20 (Financial Stability Board, 2013), which is echoed in the emergence of "need for regulatory harmonization" as a dominant topic toward the end of this window.

The positive perception spike in 2014-2015 coincides closely with topics such as the revitalization of the asset-backed securities market, long-term funding and capital markets for growth, regulation and system fragmentation, stability and financial integration, risk management and prudential measures, and market-based financing and systemic risk. This pattern aligns with Rule 5, under which sentences referencing *already implemented* regulation, as opposed to calls for future regulation, are labeled positive, since such measures are interpreted as safeguarding rather than merely flagging financial stability. The timing is consistent with the practical roll-out of the post-crisis non-banks reform agenda following the FSB's 2013 policy framework (Financial Stability Board, 2013). The shift from Rule 3-type language (need for regulation) to Rule 5-type language (regulation delivered)

is itself informative: it suggests that the ECB's negative perception in the preceding period was driven less by an inherent hostility toward non-banks than by uncertainty about whether the sector's risks were adequately contained, a concern that abates once concrete prudential measures are in place.

Perception turns negative again over 2016-2020, with topics centered on financial contagion, the transmission of monetary policy, integration and efficiency, financial stability and regulation, and risk-sharing and fiscal mechanisms. This period is captured primarily by Rule 1 and Rule 2: as non-bank intermediaries continued to grow in relative importance following the post-crisis retrenchment of bank balance sheets, ECB communication increasingly focused on whether accommodative monetary policy was being transmitted efficiently through non-bank channels, and on whether the growing interconnectedness between banks and non-banks, for instance through funding relationships or shared exposures, created new channels for the propagation of shocks. This concern is consistent with a broader empirical literature documenting that non-bank lending can substitute for bank lending in response to monetary policy changes, altering the overall transmission mechanism in ways that remain the subject of ongoing research (Anyfantaki et al., 2026, Bednarek et al., 2025, Hodula and Tiza Mimun, 2025, Tiza Mimun et al., 2025).

The small positive spike in 2021 is associated with sustainable finance, alongside continued discussion of NBFI and monetary policy transmission, systemic risk, and macroprudential measures. The sustainable finance component aligns with Rule 6: ECB speeches in this period highlight the role of non-bank intermediaries, particularly investment funds and insurers, in channeling capital toward the green transition, a contribution to real-economy financing that is coded as positive even as concerns about systemic risk and monetary transmission persist alongside it. The coexistence of positive and negative topics within the same year illustrates why our index is constructed at the sentence level: aggregate perception in any given period reflects a balance between distinct, and sometimes opposing, narratives rather than a single uniform stance.

The most recent period, 2022-2024, is dominated by negative topics: cryptocurrency and stability and regulation, financial market trends, decentralized finance (DeFi), regulatory challenges and innovation, financial stability and regulation, risk-sharing and fiscal mechanisms, NBFI sector vulnerabilities, and liquidity and

credit risks. These map onto Rule 3 (explicit calls for regulation) and Rule 2 (interconnectedness and contagion), and reflect a period in which the ECB's own financial stability analysis identified crypto-asset markets as a rapidly growing segment with increasing ties to traditional finance, linking rising systemic risk in this space to the degree of interconnectedness with the traditional financial sector as well as to growing leverage and lending activity within the crypto ecosystem (Hermans et al., 2022). The collapse of the TerraUSD stablecoin in May 2022 reinforced concerns that decentralized finance, while replicating many functions of traditional finance, remains exposed to comparable vulnerabilities without equivalent prudential safeguards, a concern that fed into the ECB's subsequent calls for the sector to be brought under effective supervision and regulation (Born and Vendrell Simón, 2022). This narrative is a continuation of the data-gap and interconnectedness concerns first identified under Rule 2 and Rule 4 in earlier periods, now applied to a structurally new but economically analogous segment of the non-bank sector.

5 Robustness

Prior to sentiment classification, we perform a series of robustness checks for the topic labeling stage. We test multiple prompt formulations and alternative model configurations to identify the most reliable specification. The final prompt is selected based on the highest classification accuracy. We further compare topic labels across different models and retain those with the highest level of agreement.

We assess the sensitivity of the index to alternative aggregation methods. Equation 2 restricts the denominator to positive and negative sentences, excluding neutral ones. Figures 6 and 7 plot both indices jointly and show that they exhibit very similar dynamics over time. Figures 8 and 9 further illustrate that the indices generated by the three best-performing models — a fine-tuned RoBERTa classifier and two fine-tuned Llama-3 variants — closely co-move.

$$\text{Perception}_{d,t} = \frac{\sum\{\text{positive sentences}_{d,t}\} - \sum\{\text{negative sentences}_{d,t}\}}{\sum\{\text{positive sentences}_{d,t}\} + \sum\{\text{negative sentences}_{d,t}\}} \quad (2)$$

In addition, we evaluated a broad set of alternative fine-tuning specifications. We

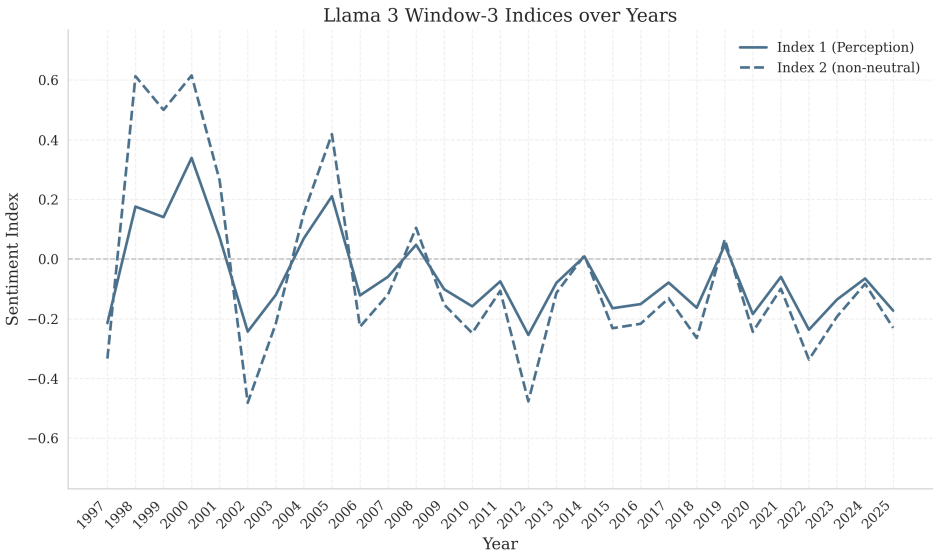


Figure 6: ECB-NBFI Perception Indices — Yearly

Note: In this figure we plot the time series for the perception index using two aggregation methods. Index 1 accounts for neutral sentences. Index 2 only accounts for positive and negative perception. Aggregated over years.

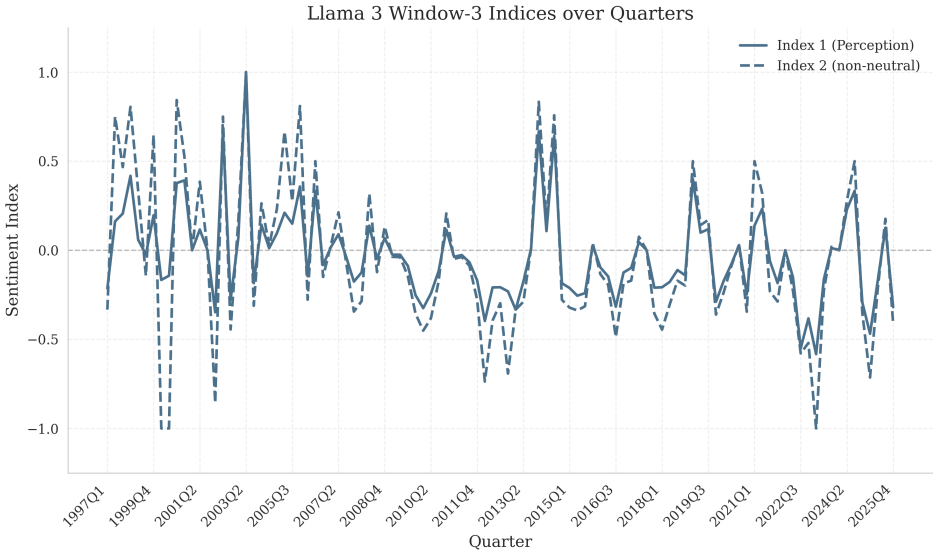


Figure 7: ECB-NBFI Perception Indices — Quarterly

Note: In this figure we plot the time series for the perception index using two aggregation methods. Index 1 accounts for neutral sentences. Index 2 only accounts for positive and negative perception. Aggregated over quarters.

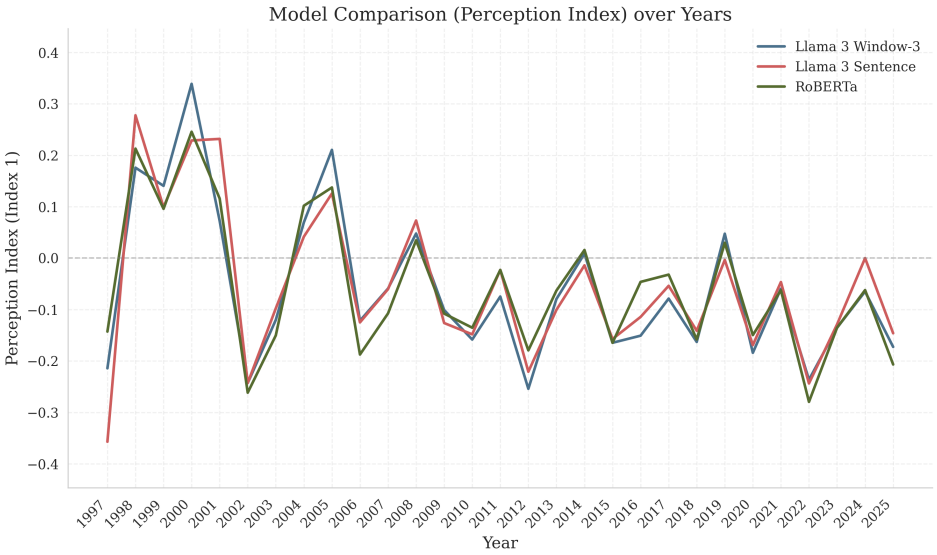


Figure 8: ECB-NBFI Perception Index — Top Performing Models, Yearly

Note: In this figure we plot the time series for the perception index using the prediction from the three top performing fine-tuned models. Aggregated over years.

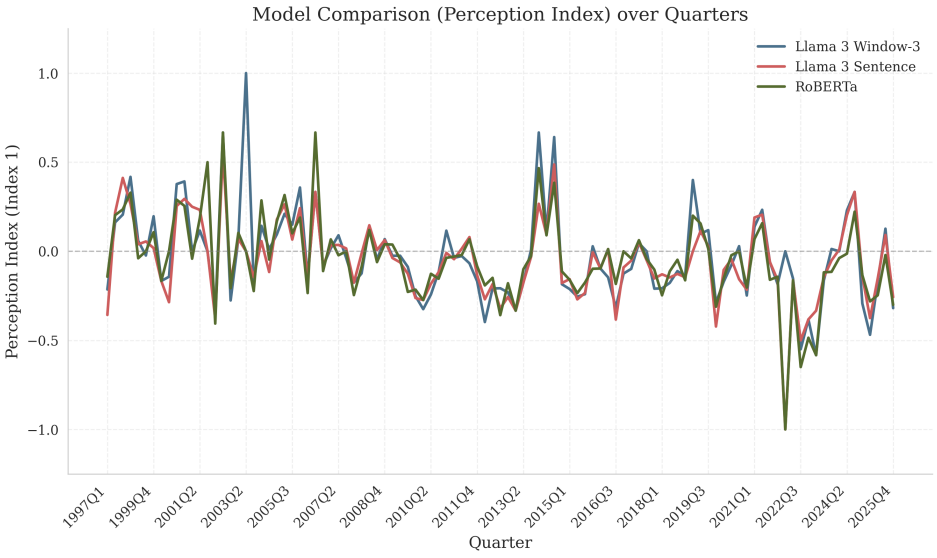


Figure 9: ECB-NBFI Perception Index — Top Performing Models, Quarterly

Note: In this figure we plot the time series for the perception index using the prediction from the three top performing fine-tuned models. Aggregated over quarters.

further fine-tuned FinBERT, RoBERTa, Llama2 in 7b-hf and in 13b-hf and Llama-3. The fine-tuning was done for the LLMs on context windows of three and eleven sentences as well as on sentence-level classification. The performance metrics reported in Table 5 in Appendix B show that the model selected for the main index consistently outperforms the alternatives in all specifications.

Finally, we address the inherent class imbalance in the training data, which reflects the uneven distribution of sentiment labels in ECB speeches. To this end, we implement up-sampling strategies for underrepresented classes (neutral and positive). As shown in Table 6 in Appendix B, these adjustments do not improve classification performance and in several cases lead to worse outcomes, consistent with overfitting to the up-sampled minority classes.

6 Conclusion

This paper develops a novel machine-learning-based index to quantify the ECB's publicly expressed perception of non-bank financial intermediaries. Drawing on a large corpus of ECB speeches spanning 1998 to 2025 and employing a large language model fine-tuned on a manually labeled dataset of ECB speeches, we document a gradual shift toward a more negative perception over time, with episodic positive deviations. The fine-tuning process involves training a pre-trained Llama3-8B model on 1,300 manually labeled sentences referring to the NBFI sector, using a context window of three sentences to capture surrounding context. The resulting model achieves an F1 score of 0.80, outperforming off-the-shelf alternatives.

Our analysis further shows that these perceptions are systematically linked to distinct narrative themes. Positive perception dominates in periods where ECB communication emphasizes the contribution of non-banks to market integration, diversification, and credit supply, as in the pre-GFC years and, more narrowly, around the 2021 discussion of sustainable finance. Negative perception, by contrast, is concentrated in periods where communication centers on the transmission of monetary policy through non-bank channels, on the interconnectedness between banks and non-banks as a potential source of contagion, and on gaps in the data available to monitor the sector, a pattern most pronounced during and after the

global financial crisis (GFC) and again in the most recent period, as concerns shift toward crypto-assets and decentralized finance.

The positive spike in 2014-2015 coincides with the rollout of the post-crisis NBFI reform agenda, a period in which ECB communication moves from calls for regulation toward references to measures already in place.

An extensive set of robustness checks confirms the stability of these findings across alternative model specifications and aggregation methods. We make the manually labeled dataset and the resulting index available to other researchers upon request, the former as training data or as a benchmark for evaluating alternative classifiers, the latter as an input for empirical work on central bank communication and non-bank finance. By providing a quantitative measure of central bank perceptions of non-bank entities, and by linking the sign of that perception to specific, identifiable economic channels, the proposed index offers a new analytical tool for studying how central bank communication about non-banks evolves, and for future work linking it to market outcomes in an increasingly market-based financial system.

References

Tobias Adrian and Hyun Song Shin. Liquidity and leverage. Journal of Financial Intermediation, 19(3):418–437, 2010. doi: 10.1016/j.jfi.2008.12.002.

Carlo Altavilla, Refet S. Gurkaynak, Luc Laeven, and Thilo Kind. Monetary transmission with frequent policy events. ECB Working Paper, (3157), 2025.

Sofia Anyfantaki, Dominic Cucic, Daniel Fricke, Philipp Hartmann, Christoph Kaufmann, Elizaveta Lukmanova, Angela Maddaloni, and Ricardo Barahona. Monetary policy transmission and non-bank financial intermediation. Technical report, ECB Occasional Paper No 391, 2026.

Mikael Apel and Marianna Blix Grimaldi. The information content of central bank minutes. Technical report, Sveriges Riksbank working paper series, 2012.

Boragan Aruoba and Thomas Drechsel. Identifying Monetary Policy Shocks: A Natural Language Approach. CEPR Discussion Papers 17133, C.E.P.R. Discussion Papers, March 2022. URL https://ideas.repec.org/p/cpr/ceprdp/17133.html.

P. Bednarek, O. Briukhova, S. Ongena, and N. Westernhagen. Effects of bank capital requirements on lending by banks and non=-bank financial institutions. Journal of Financial Intermediation, 63, 2025.

Hamza Bennani. Central bank communication in the media and investor sentiment. Journal of Economic Behavior Organization, 176:431–444, 2020. ISSN 0167-2681. doi: https://doi.org/10.1016/j.jebo.2020.05.022. URL https://www.sciencedirect.com/science/article/pii/S016726812030175X.

Hamza Bennani and Matthias Neuenkirch. The (home) bias of european central bankers: new evidence based on speeches. Applied Economics, 49(11):1114–1131, 2017. doi: 10.1080/00036846.2016.1210782. URL https://doi.org/10.1080/00036846.2016.1210782.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. Journal of Machine Learning Research, 3(4-5):993 – 1022, 2003. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-0141607824&partnerID=40&md5=505ce8839ae28d1cb56a7ff91bd0ad2d. Cited by: 31452.

Alexandra Born and Josep M. Vendrell Simón. A deep dive into crypto financial risks: Stablecoins, DeFi and climate transition risk. Macroprudential Bulletin, (18), July 2022. URL https://www.ecb.europa.

eu/press/financial-stability-publications/macroprudential-bulletin/html/ecb.mpbu202207_1~750842714e.en.html.

Benjamin Born, Michael Ehrmann, and Marcel Fratzscher. Central Bank Communication on Financial Stability. Economic Journal, 124(577):701–734, June 2014. URL https://ideas.repec.org/a/wly/econjl/v124y2014i577p701-734.html.

Jean-Charles Bricongne, Raquel Caldeira, and Baptiste Meunier. Should central banks care about text mining? A literature review. Working Paper Series 950, Banque de France, June 2024. URL https://www.banque-france.fr/en/publications-and-statistics/publications/should-central-banks-care-about-text-mining-literature-review.

Markus K. Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. Review of Financial Studies, 22(6):2201–2238, 2009. doi: 10.1093/rfs/hhn098.

Yong Cai, Santiago Camara, and Nicholas Capel. It's not always about the money, sometimes it's about sending a message: Evidence of Informational Content in Monetary Policy Announcements. Papers 2111.06365, arXiv.org, November 2021. URL https://ideas.repec.org/p/arx/papers/2111.06365.html.

Olivier Coibion, Yuriy Gorodnichenko, and Saten Kumar. How Do Firms Form Their Expectations? New Survey Evidence. American Economic Review, 108(9):2671–2713, September 2018. URL https://ideas.repec.org/a/aea/aecrev/v108y2018i9p2671-2713.html.

Ricardo Correa, Keshav Garud, Juan M Londono, and Nathan Mislang. Sentiment in central banks' financial stability reports. Review of Finance, 25(1):85–120, 2021.

L. Dräger and G. Nghiem. Inflation Literacy, Inflation Expectations, and Trust in the Central Bank: A Survey Experiment. CESifo Working Paper, 10539, 2023.

Michael Ehrmann and Marcel Fratzscher. The timing of central bank communication. European Journal of Political Economy, 23(1):124–145, March 2007. URL https://ideas.repec.org/a/eee/poleco/v23y2007i1p124-145.html.

Michael Ehrmann and Jonathan Talmi. Starting from a blank page? Semantic similarity in central bank communication and market volatility. Journal of Monetary Economics, 111(C):48–62, 2020. doi: 10.1016/j.jmoneco.2019.01. URL https://ideas.repec.org/a/eee/moneco/v111y2020icp48-62.html.

Financial Stability Board. Policy framework for strengthening oversight and regulation of shadow banking entities. Technical report, Financial Stability Board, August 2013. URL https://www.fsb.org/2013/08/r_130829c/.

Eric Fischer, Rebecca McCaughrin, Saketh Prazad, and Mark Vandergon. Fed Transparency and Policy Expectation Errors: A Text Analysis Approach. Staff Reports 1081, Federal Reserve Bank of New York, November 2023. URL https://ideas.repec.org/p/fip/fednsr/97356.html.

E. Franceschi, M. Grodzicki, B. Kagerer, C. Kaufmann, F. Lenoci, L. Mingarelli, C. Pancaro, and R. Senner. Key linkages between banks and the non-bank financial sector. Financial Stability Review, 1, 2023.

Richard Frankel, Jared Jennings, and Joshua Lee. Disclosure sentiment: Machine learning vs. dictionary methods. Management Science, 68(7):5514–5532, 2022. URL https://EconPapers.repec.org/RePEc:inm:ormnsc:v:68:y:2022:i:7:p:5514-5532.

Margherita Giuzio, Christoph Kaufmann, Ellen Ryan, and Lorenzo Cappiello. Investment funds, risk taking, and monetary policy in the euro area. ECB Working Paper, (2605), 2021.

Roberto Gómez-Cram and Marco Grotteria. Real-time price discovery via verbal communication: Method and application to Fedspeak. Journal of Financial Economics, 143(3):993–1025, 2022. doi: 10.1016/j.jfineco.2021.12. URL https://ideas.repec.org/a/eee/jfinec/v143y2022i3p993-1025.html.

Refet S. Gürkaynak, Brian Sack, and Eric Swanson. The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models. American Economic Review, 95(1):425–436, March 2005. URL https://ideas.repec.org/a/aea/aecrev/v95y2005i1p425-436.html.

Andrew Haldane, Alistair Macaulay, and Michael McMahon. The 3 E's of central bank communication with the public. Bank of England working papers 847, Bank of England, 2020. URL https://ideas.repec.org/p/boe/boeewp/0847.html.

Lieven Hermans, Annalaura Ianiro, Urszula Kochanska, Veli-Matti Törmälehto, Anton van der Kraaij, and Josep M. Vendrell Simón. Decrypting financial stability risks in crypto-asset markets. Financial Stability Review, 1, 2022. URL https://www.ecb.europa.eu/press/financial-stability-publications/fsr/special/html/ecb.fsrart202205_02~1cc6b111b4.en.html. Special Feature A, May 2022.

Martin Hodula and Anisa Tiza Mimun. Investment funds and the monetary-macroprudential policy interplay. ECB Working Paper, (3151), 2025.

Federic Holm-Hadulla, Falk Mazelis, and Sebastian Rast. Bank and non-bank balance sheet responses to monetary policy shocks. Economic Letters, 222(110918), 2023.

Feiming Huang, Lei Chen, Wei Guo, Xianchao Zhou, Kaiyan Feng, Tao Huang, and Yudong Cai. Identifying covid-19 severity-related sars-cov-2 mutation using a machine learning method. Life, 12(6), 2022. ISSN 2075-1729. doi: 10.3390/life12060806. URL https://www.mdpi.com/2075-1729/12/6/806.

Xisong Jin and Francisco Nadal De Simone. Monetary policy and systemix risk-taking in the euro area investment fund industry: A structural factor-augmented vector autoregression analysis. Journal of Financial Stability, 49(100749), 2020.

Rokas Kaminskas and Linas Jurkaias. Ecb communication sentiments: How do they relate to the economic environment and financial markets? Journal of Economics and Business, 131(C):None, None 2024. doi: 10.1016/j.jeconbus.2024.106198.

Christoph Kaufmann. Investment funds, monetary policy, and the global financial cycle. Journal of the European Economic Association, 21:593–636, 2023.

Tim Loughran. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. The Journal of Finance, 66:35 – 65, 02 2011. doi: 10.1111/j.1540-6261.2010.01625.x.

Haroon Mumtaz, Jumana Saleheen, and Roxane Spitznagel. Keep it Simple: Central Bank Communication and Asset Prices. Working Papers 960, Queen Mary University of London, School of Economics and Finance, July 2023. URL https://ideas.repec.org/p/qmw/qmwecw/960.html.

Tim Munday and James Brookes. Mark my words: the transmission of central bank communication to the general public via the print media. Bank of England working papers 944, Bank of England, October 2021. URL https://ideas.repec.org/p/boe/boeewp/0944.html.

Giang Nghiem, Lena Dräger, and Ami Dalloul. Anchoring Households' Inflation Expectations when Inflation is High. Hannover Economic Papers (HEP) dp-719, Leibniz Universität Hannover, Wirtschaftswissenschaftliche Fakultät, October 2024. URL https://ideas.repec.org/p/han/dpaper/dp-719.html.

Conor Parle. The financial market impact of ECB monetary policy press conferences — a text based approach. European Journal of Political Economy, 74: 102230, 2022. doi: 10.1016/j.ejpoleco.2022.102230.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In Marilyn Walker, Heng Ji, and Amanda Stent, editors, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1202. URL https://aclanthology.org/N18-1202/.

Moritz Pfeifer and Vincent P. Marohl. Centralbankroberta: A fine-tuned large language model for central bank communications. The Journal of Finance and Data Science, 9:100114, 2023. ISSN 2405-9188. doi: https://doi.org/10.1016/j.jfds.2023.100114. URL https://www.sciencedirect.com/science/article/pii/S2405918823000302.

Matthieu Picault and Thomas Renault. Words are not all created equal: A new measure of ecb communication. Journal of International Money and Finance, 79:136–156, 2017. ISSN 0261-5606. doi: https://doi.org/10.1016/j.jimonfin.2017.09.005. URL https://www.sciencedirect.com/science/article/pii/S0261560617301808.

Lynnette Purda and David Skillicorn. Accounting Variables, Deception, and a Bag of Words: Assessing the Tools of Fraud Detection. Contemporary Accounting Research, 32(3):1193–1223, September 2015. doi: 10.1111/1911-3846.12089. URL https://ideas.repec.org/a/wly/coacre/v32y2015i3p1193-1223.html.

Alec Radford. Improving language understanding by generative pre-training. 2018.

Maik Schmeling and Christian Wagner. Does central bank tone move asset prices? Journal of Financial and Quantitative Analysis, page 1–32, 2024. doi: 10.1017/S0022109024000073.

Herman Stekler and Hilary Symington. Evaluating qualitative forecasts: The fomc minutes, 2006–2010. International Journal of Forecasting, 32(2):559–570, 2016. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2015.02.003. URL https://www.sciencedirect.com/science/article/pii/S0169207015000588.

Anisa Tiza Mimun, Gabor Fukker, and Matthias Sydow. The effects of monetary policy on banks and non-banks in times of stress. ECB Working Paper, (3114), 2025.

A Fine-tuning for sentiment- Specifications

Model	GAS	Batches	Epochs
FinBert	-	16	5
RoBERTa	-	16	5
Llama2-7-hf-Sentence	1	4	3
Llama2-7-hf-Window-3	1	4	3
Llama2-7-hf-Window-11	1	4	3
Llama2-13b-hf-Sentence	1	4	3
Llama2-13b-hf-Window-3	1	4	3
Llama2-13b-hf-Window-11	1	4	3
Llama3-8B-Sentence	1	4	3
Llama3-8B-Window-3	1	4	3
Llama3-8B-Window-11	1	4	3

Table 3: Hyperparameters for classifier model

B Robustness- other model specifications

Model	Precision	Recall	F1
RoBERTa Baseline	0.06	0.24	0.10
RoBERTa fine-tuned	0.70	0.71	0.70
FinBert Baseline	0.31	0.30	0.25
FinBert fine-tuned	0.69	0.67	0.69
Llama2-7-hf Baseline	0.28	0.33	0.20
Llama2-7-hf fine-tuned	0.65	0.62	0.60
Llama2-13b-hf Baseline	0.33	0.36	0.24
Llama2-13b-hf fine-tuned	0.63	0.60	0.58

Table 4: Results of classifiers before and after fine-tuning

Note: Precision, recall, and F1 are weighted averages across the three perception classes, computed on the held-out test set. The RoBERTa baseline is based on `roberta-base`, a masked language model without a trained classification head; the reported figures (for RoBERTa) therefore reflect an untrained classifier and are included for completeness rather than as a measure of zero-shot performance.

Model	Precision	Recall	F1
FinBERT Baseline	0.29	0.32	0.27
FinBERT fine-tuned	0.61	0.62	0.61
RoBERTa Baseline	0.06	0.24	0.10
RoBERTa fine-tuned	0.70	0.71	0.70
Llama2-7-hf Baseline-s.	0.56	0.34	0.31
Llama2-7-hf fine-tuned-s.	0.66	0.64	0.63
Llama2-13b-hf Baseline-s.	0.20	0.36	0.26
Llama2-13b-hf fine-tuned-s.	0.69	0.69	0.69
Llama3-8B Baseline-s.	0.52	0.32	0.26
Llama3-8B fine-tuned-s.	0.69	0.70	0.70
Llama2-7-hf Baseline-w3	0.57	0.27	0.20
Llama2-7-hf fine-tuned-w3	0.50	0.54	0.51
Llama2-13b-hf Baseline-w3	0.11	0.26	0.12
Llama2-13b-hf fine-tuned-w3	0.35	0.28	0.22
Llama3-8B Baseline-w3	0.52	0.26	0.19
Llama3-8B fine-tuned-w3	0.79	0.80	0.80
Llama2-7-hf Baseline-w11	0.67	0.22	0.19
Llama2-7-hf fine-tuned-w11	0.53	0.51	0.52
Llama2-13b-hf Baseline-w11	0.14	0.32	0.19
Llama2-13b-hf fine-tuned-w11	0.40	0.33	0.30
Llama3-8B Baseline-w11	0.67	0.26	0.18
Llama3-8B fine-tuned-w11	0.73	0.73	0.73

Table 5: Results of classifiers before and after fine-tuning all specification

Note: Precision, recall, and F1 are weighted averages across the three perception classes, computed on the held-out test set. The RoBERTa baseline is based on `roberta-base`, a masked language model without a trained classification head; the reported figures (for RoBERTa) therefore reflect an untrained classifier and are included for completeness rather than as a measure of zero-shot performance.

Model	Precision	Recall	F1
FinBERT Baseline	0.25	0.24	0.24
FinBERT fine-tuned	0.59	0.60	0.58
RoBERTa Baseline	0.06	0.24	0.10
RoBERTa fine-tuned	0.65	0.66	0.65
Llama2-7-hf Baseline-s.	0.56	0.34	0.31
Llama2-7-hf fine-tuned-s.	0.20	0.36	0.26
Llama2-13b-hf Baseline-s.	0.20	0.36	0.26
Llama2-13b-hf fine-tuned-s.	0.73	0.67	0.68
Llama3-8B Baseline-s.	0.52	0.33	0.28
Llama3-8B fine-tuned-s.	0.70	0.69	0.69
Llama2-7-hf Baseline-w3	0.49	0.26	0.15
Llama2-7-hf fine-tuned-w3	0.65	0.62	0.62
Llama2-13b-hf Baseline-w3	0.60	0.34	0.20
Llama2-13b-hf fine-tuned-w3	0.73	0.39	0.28
Llama3-8B Baseline-w3	0.65	0.27	0.16
Llama3-8B fine-tuned-w3	0.75	0.75	0.74
Llama2-7-hf Baseline-w11	0.21	0.26	0.15
Llama2-7-hf fine-tuned-w11	0.45	0.40	0.40
Llama2-13b-hf Baseline-w11	0.12	0.35	0.18
Llama2-13b-hf fine-tuned-w11	0.45	0.44	0.43
Llama3-8B Baseline-w11	0.21	0.32	0.22
Llama3-8B fine-tuned-w11	0.65	0.63	0.62

Table 6: Results of classifiers before and after fine-tuning with upsampling

Note: Precision, recall, and F1 are weighted averages across the three perception classes, computed on the held-out test set. The RoBERTa baseline is based on `roberta-base`, a masked language model without a trained classification head; the reported figures (for RoBERTa) therefore reflect an untrained classifier and are included for completeness rather than as a measure of zero-shot performance.

Acknowledgements

We thank Lena Dräger, Loukas Karabarbounis, Michele Lenza, Moritz Pfeifer, and Peter Tillmann for their helpful comments, questions and discussions. We are also thankful to DGR brownbag research seminar participants as well as PIEC conference participants for their helpful feedback. AI has been used for redrafting the text throughout the paper.
The views expressed in the paper are those of the authors and do not represent the ECB or the Eurosystem.

Ami Dalloul
University Duisburg-Essen, Essen, Germany; email: ami.dalloul@wiwi.uni-due.de

Anisa Tiza Mimun
European Central Bank, Frankfurt am Main, Germany; email: anisa.tiza_mimun@ecb.europa.eu

© European Central Bank, 2026

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

This paper can be downloaded without charge from www.ecb.europa.eu, from the Social Science Research Network electronic library or from RePEc: Research Papers in Economics. Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website.

PDF ISBN 978-92-899-8071-5 ISSN 1725-2806 doi:10.2866/3302263 QB-01-26-248-EN-N