ANALYSIS OF CBDC NARRATIVE OF CENTRAL BANKS USING LARGE LANGUAGE MODELS

2023

BANCO DE ESPAÑA
Eurosistema

Documentos de Trabajo
N.º 2321

Andres Alonso-Robisco and José Manuel Carbó

ANALYSIS OF CBDC NARRATIVE OF CENTRAL BANKS USING LARGE LANGUAGE MODELS

ANALYSIS OF CBDC NARRATIVE OF CENTRAL BANKS USING LARGE LANGUAGE MODELS [*]

Andres Alonso-Robisco

BANCO DE ESPAÑA

José Manuel Carbó

BANCO DE ESPAÑA

(*) The authors work as senior economists in the Financial Innovation Division of Banco de España, and appreciate the comments received from José Manuel Marqués, Ana Fernández, Sergio Gorjón and José Luis Romero. The opinions and analyses expressed in this paper are the responsibility of the authors and, therefore, do not necessarily match with those of the Banco de España or the Eurosystem.

Documentos de Trabajo. N.º 2321

August 2023

https://doi.org/10.53479/33412

ISSN: 1579-8666 (on line)

Abstract

Central banks are increasingly using verbal communication for policymaking, focusing not only on traditional monetary policy, but also on a broad set of topics. One such topic is central bank digital currency (CBDC), which is attracting attention from the international community. The complex nature of this project means that it must be carefully designed to avoid unintended consequences, such as financial instability. We propose the use of different Natural Language Processing (NLP) techniques to better understand central banks' stance towards CBDC, analyzing a set of central bank discourses from 2016 to 2022. We do this using traditional techniques, such as dictionary-based methods, and two large language models (LLMs), namely Bert and ChatGPT, concluding that LLMs better reflect the stance identified by human experts. In particular, we observe that ChatGPT exhibits a higher degree of alignment because it can capture subtler information than BERT. Our study suggests that LLMs are an effective tool to improve sentiment measurements for policy-specific texts, though they are not infallible and may be subject to new risks, like higher sensitivity to the length of texts, and prompt engineering.

Keywords: ChatGPT, BERT, CBDC, digital money.

JEL classification: G15, G41, E58.

Resumen

Los bancos centrales están utilizando cada vez más la comunicación verbal en su estrategia, abarcando no solo la política monetaria tradicional, sino también un amplio conjunto de temas. Uno de estos temas es la moneda digital de los bancos centrales (CBDC, por sus siglas en inglés), que está captando la atención de la comunidad internacional. La naturaleza compleja de este proyecto implica que debe ser diseñado cuidadosamente para evitar consecuencias no deseadas, como la inestabilidad financiera. En este trabajo, proponemos utilizar diferentes técnicas de Procesamiento de Lenguaje Natural (NLP, por sus siglas en inglés) para comprender mejor la postura o sentimiento de los bancos centrales hacia el CBDC, analizando un conjunto de discursos de los bancos centrales desde 2016 hasta 2022. Para ello, utilizamos técnicas tradicionales, como los métodos basados en diccionarios, y dos Modelos de Lenguaje de Gran Tamaño (LLM, por sus siglas en inglés) como BERT y ChatGPT, llegando a la conclusión de que los LLM reflejan mejor la postura identificada por los expertos humanos. En particular, observamos que ChatGPT muestra un mayor grado de alineación porque puede capturar información más sutil que BERT. Nuestro estudio sugiere que los LLM son una herramienta eficaz para mejorar las mediciones de sentimiento en textos específicos de contenido estratégico, aunque no son infalibles y pueden estar sujetos a nuevos riesgos, como una mayor sensibilidad a la longitud de los textos y el diseño de las preguntas realizadas al propio modelo.

Palabras clave: ChatGPT, BERT, CBDC, moneda digital.

Códigos JEL: G15, G41, E58.

1. Introduction

Central banks play a fundamental role in the modern economy. One of the channels through which they exert their influence is through communication, by affecting financial markets expectations to make monetary policy more predictable (McKay et al. 2016). For example, they can communicate on future policy intentions, increase transparency in their decision-making processes, or publishing economic projections. A well-known example is to steer market expectations using "forward-guidance", a tool employed by some central banks to communicate their intentions about the future path of key policy variables, such as interest rate, although there exist limitations to its power (Campbell et al. 2019; Cole 2021; McKay et al. 2016). Another example is the publication of inflation forecasts, which can influence the shape of yield curves (Hansen et al. 2019), households' inflation expectations through "open mouth operations" (Guthrie and Wright 2000), or written communication from supervisory bodies outlining corrective measures (Goldsmith-Pinkham et al. 2016). Overall, it seems that if central banks' communication is done effectively, it builds credibility and maintains a strong reputation for their commitment with the general public (Bholat et al. 2019; Blinder et al. 2008; Blinder 2018; Haldane and McMahon 2018).

Interestingly, beyond monetary policy and supervisory actions, in recent times, central banks are expanding the scope of their communication. A topic that is gaining momentum within central banks communication is Central Bank Digital Currency (CBDC), a new form of money that exists only in digital form. The implementation of a CBDC could enable central banks to engage in large-scale intermediation for retail and/ or wholesale deposits (Fernandez-Villaverde et al. 2021). As we will examine further, several aspects of a CBDC remain in discussion. Therefore, for the sake of simplicity, we will adhere to Barrdear and Kumhof (2016) definition, which describes it as providing electronic access to domestic currency denominated balance-sheets of central banks, independent of the technological solution applied (whether token-based or account-based), or any other architecture details (e.g.: interest-bearing deposits or not). Depending on the final design, a CBDC could serve various purposes, such as payment system integrity, promoting financial inclusion (Auer et al. 2022a), and fostering innovation, among others. On the other hand, its introduction could suppose unwanted effects on the financial system, such as a flight from commercial deposits, destabilizing financial intermediation, anonymity issues or privacy concerns (Auer and Boehme 2020; Auer et al. 2022b; Davoodalhosseini 2022; Ferrari Minesso et al. 2022).

Currently, many central banks are exploring the potential value of some sort of CBDC, though at various paces. Some are simply incorporating this topic into speeches, others are making project announcements, while a few are conducting pilots and live experimentation (Auer et al. 2020). The gaining momentum of these discussions between central banks could be attributed to the growing demand to cater to the digital economy, react to private initiatives (like Facebook's project Libra), keep pace with other central banks innovative projects, or to adapt the payment systems to an ongoing disruptive transformation, to name a few. As digital payments gradually replace cash transactions (Tarlin 2021), big tech companies compete for dominance in payments services, and new types of digital assets threatens the stability of the financial system, the relevance of CBDCs has increased Lagarde, 2022.

There are several approaches to the design of CBDCs, and these vary significantly between jurisdictions, which gives an idea of the complexity of this issue. Auer and Boehme (2020) describe the main design options, identifying the possible trade-offs in each one. The first decision to make is which operational role of the central bank might take. Options range from a role as a direct CBDC provider, to intermediary solutions

* The authors work as senior economists in the Financial Innovation Division of Banco de España, and appreciate the comments received from José Manuel Marqués, Ana Fernández, Sergio Gorjón and José Luis Romero. The opinions and analyses expressed in this paper are the responsibility of the authors and, therefore, do not necessarily match with those of the Banco de España or the Eurosystem.

where central banks maintain a central ledger of retail transactions, to solutions where only wholesale transactions are considered. A second technical design decision is the infrastructure, which could be based on Distributed Ledger Technology (DLT) or a conventional centralized database. Additionally, more dimensions of this problem concerns how will consumer access the CBDC (account-based or digital tokens) or the interlinkages of CBDC with cross-border payments.

All these possibilities in the choice of CBDC design and operation generate a high level of uncertainty around its eventual form. For this reason, we believe that it is relevant to analyze the tone of the central banks (represented by speeches by their governors) to enhance the predictability and transparency at this time when several of the CBDC projects are still in the initial stages of development. A better measurement of the sentiment towards CBDC could refine central banks' intentions and policy direction, as well as improve its perceived credibility and help managing market expectations and investor sentiment (Scharnowski 2022).

To this purpose, in this paper we aim to quantify central banks' sentiment towards CBDC using two of the most widely used Large Language Models (LLMs) such as ChatGPT (OpenAI, 2023), and BERT (Devlin et al. 2019), in particular, a pre-trained version of BERT for financial communication known as FinBERT (Yang et al. 2020).[1] We do this by evaluating a series of central bank speeches on the topic of CBDC, and we compare the sentiment obtained with LLMs with that obtained through a dictionary-based method (i.e.: polarity), and with the sentiment from human experts or labeled data. Our contribution is twofold. First, to the best of our knowledge, we harness for the first time the power of recent developments in LLMs to compute sentiment in an area highly relevant area to financial regulation like digital money. Second, by analyzing the sentiment calculated by each method, we can understand the potential benefits and limitations of each approach. The main conclusions are that the sentiment towards CBDC expressed by central banks is increasingly positive, and that the sentiment obtained with ChatGPT is the most reliable, since it is the most similar to the labeled data because it is able to capture the subtleties of the text and is less noisy. To this purpose, we use the recently updated (January, 2023) database of CBDC speeches and reports by central banks compiled by Auer et al. (2020). This database offers a unique opportunity to compare the results of language models to the expert labels that humans assign to each speech stance.

The remaining part of this paper is structured as follows. Section 2 details the relevant literature review to give pertinent context to this study. Section 3 explains the dataset, Section 4 delves deeper into the methodology of all three language models, Section 5 provides the results, and Section 6 concludes, with remarks for further research.

2. Literature review

As mentioned before, many central banks are contemplating whether to issue a CBDC, as it has potential benefits, as an attractive public alternative to traditional deposits held in private banks (Fernandez-Villaverde et al. 2021). Interestingly, Adrian and Mancini-Griffoli (2021) provide a framework for comparing traditional forms of money with their new digital equivalent. However, using a CBDC might also be costly for agents having therefore implications in terms of financial stability and the design of an optimal monetary policy strategy (Davoodalhosseini 2022; Ferrari Minesso et al. 2022). To deepen on the motivation and design features CBDC projects worldwide we revert to Auer et al. (2022b) which offers a guided tour of the growing CBDC literature on the microeconomic considerations related to projects architectures, technologies, and privacy as well as the macroeconomic implications for the economy and

[1] From now on we will use the generic term BERT, though referring always to this particular class.

the financial system. For instance, following this line of research, Siklos et al. (2018) study the macroeconomic impact of CBDC, concluding that its adoption needs not to impair inflation control, and Barrdear and Kumhof (2016) find a potential positive impact on Gross Domestic Product in the US. On the other hand, (Andolfatto 2020) investigates important microeconomics issues, like how CBDC can be expected to impact a monopolistic banking sector, concluding that a properly designed CBDC is not likely to threaten financial stability. Indeed, there is still an ample range of uncertainty surrounding the impact of CBDC depending on its final design attributes (e.g.: anonymity, rate of return, etc.).[2] This translates in wide estimates for its adoption. For instance, Li (2023) finds that Canadian households' demand for CBDC can range from 4% to 52% of total liquid assets, and banks endogenous responses to CBDC could considerably constraint its upper bound take-up to 20%. In parallel, Kiff et al. (2020), is a valuable survey that reviews the processes, roles, and responsibilities that would need to be defined for the issuance of a retail CBDC, while Gorjón Rivas (2022), and Romero Romero Ugarte et al. (2021) offers some insights on wholesale CBDC based on different technologies, like distributed ledgers (DLT).

A critical article for understanding the increased interest in CBDCs among the central banking community is Auer et al. (2020). In that research, the authors take stock of design efforts, technical approaches, and differing stances on CBDC issuance across jurisdictions that have conducted CBDC experimentation. To do this, from a database of more than 16,000 central bank speeches by central bank representatives, they extract 351 texts that explicitly mention CBDC. In addition, they evaluate with their expert judgment the political position of these speeches.[3] Therefore, as our goal is to measure the sentiment of central banks towards CBDC, we will compare the performance of Large Language Models (LLMs) with traditional NLP techniques (like dictionary-based methods), and on a novel way, with human expert labels. This will allow to assess the benefits and limitations of these models.

Firstly, we find that studying the sentiment of central bank communication is a relevant research topic, as highlighted in studies such as (Born et al. 2014), who show how sentiment of central banks' speeches about financial stability have a significant effect on market returns and volatility. Also Hansen et al. (2018) study central bank transparency using topic modeling in an event study around 1993, the year the Fed started to release the FOMC meeting transcripts. Notably, the evolution of NLP is accelerating the research in this area. A good example is Hansen and Kazinnik (2023), who use GPT models to decipher the communication from the Federal Reserve, finding that these novel models surpass the performance of more traditional classification methods.

On the other hand, the magnitude of the potential implications for an open-wide economy of the introduction of a CBDC is reflected in Ferrari Minesso et al. (2022) who analyze the international transmission of monetary policy decisions and technology shocks in the presence and absence of a CBDC. Following this line, the impact of CBDC communication in the real economy and financial markets is studied in Burlon et al. (2022), who provide evidence on the estimated effects of CBDC news on bank valuations and lending in the euro area, finding that it depends on the quantity of CBDC in circulation. Also, Wang et al. (2022), find that CBDC attention and uncertainty indices from financial news have a positive relationship on cryptocurrencies, foreign exchange, and bond markets, as well as VIX and gold. Closer to our study, Scharnowski (2022) use the raw CBDC sentiment index, as provided by Auer et al. (2020), to study the market reaction to central banks' speeches on CBDC, observing that cryptocurrency prices increase more strongly after speeches taking a positive tone. Similarly, Tian et al. (2023) relies al well on the texts compiled by Auer et al. (2020) but they propose to measure central banks' sentiment toward CBDC based on

[2] See for instance, Conlon et al. (2022) about "to CBDC or not CBDC".

[3] At the moment of drafting this paper, it is updated up to January 2023.

the polarity of the words using the Loughran-McDonald Financial Dictionary (LMFD) (Loughran and McDonald 2011). They find that different cybersecurity risks (cyberware and cyberattacks) can have an impact on the posture of central banks towards CBDCs. Specifically, they find that this posture improves as a result of losses due to cyber-attacks. They conclude that central banks believe CBDC could be a public solution to protect consumers from risks arising from the private sector.

Our study complements both literature strands, first on CBDC, and second, on central banks' communication, by means of using new techniques of (NLP to assess the sentiment of central banks' speeches in a topic that is gaining traction and relevance such as digital money Hansson (2021). As a result, the application of LLM techniques, such as ChatGPT (OpenAI, 2023), allows us to obtain a reliable sentiment measure, as a continuous variable between -1 and 1 (as opposed to the categorical judgment of experts), less noisy than Dictionary-based methods, and that captures subtler information compared to models like BERT.

2.1. *Dataset*

The two most common ways of central bank communication are: (i) speeches, and (ii) thematic reports. As a consequence of the more transparent manner that central banks are communicating, the number of speeches available has increased (Hansson 2021). In this paper we use a collection of relevant central bank speeches touching upon CBDC, collected in Auer et al. (2020). This database includes different issues related to CBDCs, such as the different motivations of central banks for their experimentation, or the public interest on this topic captured by web searches, or the architectures and design attributes of each CBDC project. More importantly to our work, it also analyzes the sentiment of those speeches using human expert knowledge. In particular, the tone of the texts has been labeled by economists from the Bank of International Settlements (BIS).[4]

As mentioned before, this database of documents has been used in different recent academic papers like Scharnowski (2022), or Tian et al. (2023). In we provide a summary of the corpus of documents available. It consist of 351 texts, from 44 different geographical areas, published during the period 2016-2022.[5] The central banks with more than 10 appearances are France, Euro Area, Germany, UK, US, Japan, Hong Kong, and Singapore, followed by 36 countries with lower representation.[6] It can be seen as well that the number of speeches on CBDC by central banks has been growing over the years, which reflects the increasing attention that CBDC has been gaining from 2016 onwards.

3. Methodology

Text data is growing, and currently it represents the vast majority of data available for researchers (Gentzkow et al. 2019; Ash and Hansen 2023). The unstructured nature of this kind of data requires special techniques to deal with it, known as NLP). These

[4]This document provides an accompanying Excel file with a collection of links to relevant CBDC speeches, ranked in temporary order, with a timestamp. The title, the speech stance (-1,0,1), and the url are listed as well, therefore we have automated a procedure to extract, pre-process and storage the text in a separate database.

[5]Importantly, while the original publications dates back to 2020, the speeches database has been updated since then until January 2023. Documents include a timestamp that allows later on for the construction of the sentiment index as a time series.

[6]The other countries include "Finland", "Canada", "Sweden", "Malaysia", "Czech Republic", "Chile", "Denmark", "Barbados", "Australia", "Pakistan", "Thailand", "Lithuania", "Italy", "Greece", "New Zealand", "Hong Kong SAR", "Belgium", "Curaçao and Saint Maarten", "Norway", "South Africa", "Kuwait", "United Arab Emirates", "China", "Bosnia and Herzegovina", "The Bahamas", "Mexico", "France", "Ireland", "Mauritius", "Switzerland", "Korea", "Spain", "Philippines", "Russia", "India", "Albania", "Netherlands", "Macedonia, the Former Yugoslav Republic of", "Indonesia", "Korea, Republic of", "Israel", "Serbia".

Table 1. Descriptive statistics of corpus of documents: cross-sectional dispersion of countries experimenting with CBDC, and the timeline of speeches collected in Auer et al. (2020).

Country	Documents
France	51
Euro Area	41
Germany	31
United Kingdom	24
United States	24
Singapore	14
Hong Kong	13
Japan	13
Italy	13
Thailand	9
Canada	9
Others	109

Year	Documents
2016	9
2017	16
2018	46
2019	42
2020	54
2021	76
2022	108

techniques are useful for different tasks to analyze text, like realizing summaries, predicting words in sentences, finding keywords or topics, or analyzing the sentiment of a given extract.

In this study we are interested on investigating the use of LLMs strictly applied to measuring the tone, stance or sentiment of a text. In this sense, we can frame LLMs as an innovation within the field of NLP. In the past 5 years the explosion of LLMs has been largely driven by the use of Transformers, as a particular architecture of deep neural network which relies on the concept of self-attention (Vaswani et al. 2017) to embed the context during its learning process.[7]

In this paper we use three techniques to measure sentiment. One of the techniques is dictionary-based (also known as bag-of-words approach), which can be considered as a benchmark as it is a traditional NLP technique. The other are two of the most widely used LLMs currently in the academic literature and industries. One the one hand, ChatGPT, a version of the general GPT (Generative Pre-trained Transformer) models, is an auto-regressive Transformer model, recently released by OpenAI. On the other hand, BERT (Bi-directional Encoder Representations from Transformers), which was released by Google experts (Devlin et al. 2019). We selected these two families of Transformers because both have strengths for sentiment analysis. BERT has the advantage of including bi-directional attention (in contrast to left, auto-regressive, attention of GPT models), while ChatGPT has been trained with a gigantic amount 45 terabytes, compared to the 3 terabytes of BERT. Our work will help to understand the strengths and limitations of both models and to evaluate the improvement that they represent with respect to dictionary-based methods when it comes to measuring sentiment.

3.1. *Dictionary-based method*

One of the most traditional and popular techniques to compute the sentiment of a text are the known as dictionary-based methods, which rely on a dictionary or predefined collection of words with positive and negative tone. There are different dictionaries or collections, depending on the nature of the texts at hand. In finance, one of the most used dictionaries is (Loughran and McDonald 2011). It contains 354 positive words and 2,337 negative words. While it is easy to use it straightforward, this approach makes strong assumptions about the meaning of specific words that were selected, so

[7] See Amatriain (2023) for further references on the evolution of Transformer models.

misspecification can cause severe noise in dictionary-based sentiment indices.[8] This concern is presented by Hayo and Zahner (2023), who perform an analysis of variance (ANOVA) to show that about 80% of the variation in sentiment in Fed and ECB speeches is due to noise, which raises questions about the index's reliability as an indicator.

In any case, due to its popularity, we find it a suitable benchmark for our exercise. Since we can count the number of positive and the negative words in the corpus of text that match Loughran and McDonald (2011), an easy way to calculate the sentiment of a text is computing the polarity. This index is a measure between -1 and 1 that is computed as follows:

$$Polarity = \frac{Positives - Negatives}{Positives + Negatives}$$

3.2. *BERT*

As mentioned before, BERT is a deep learning architecture for NLP analysis developed by Google which harness the notion of self-attention to weigh-in the context in the computation of word embedding. It was developed in 2018, with the goal of provide contextual understanding on unlabeled data (Devlin et al. 2019). The original task was to learn to predict words that might appear before and after other text, hence the term bidirectional, with the aim of translating sentences from German into English. Unlike previous LLMs relying on Feed-Forward Neural Networks and Convolutional Neural Networks, Transformers do not process the input sequence sequentially over time. Instead, they process the entire sentence at once weighting the relevance of words based on the self-attention mechanism (Vaswani et al. 2017).

In particular, for this paper, we use the version known as FinBERT (Yang et al. 2020), which is a financial domain-specific language model based on BERT, pre-trained using a large scale of financial communication corpora.[9] The sentiment output for each analyzed piece of text is a probability distribution over the two possible classes, negative and positive. Is therefore possible to transform it to the range -1 and 1 with a transformation such as:

$$Sentiment = Prob\,(positive) - Prob\,(negative)$$

3.3. *ChatGPT*

Similar to BERT, GPT models were initially trained to predict the next word in a sentence given the previous words, which allows it to learn the patterns of natural language. The main difference between them is the type of attention mechanism, as GPT models are auto-regressive, while the BERT family is bi-directional. Though, on top of this difference, it shall be noted that ChatGPT, a particular version of GPT, has been trained on a much larger corpus than any other model up to 2023. It also has been trained using human labeling, which positions this model as potential market leader (Bubeck et al. 2023). Indeed, it has gained immense popularity among early adopters and in particular between finance experts and researchers (Dowling and Lucey 2023; Hansen and Kazinnik 2023).[10]

To interact with ChatGPT it is necessary to specify the task to be carried out by means of a prompt. The prompt can be as flexible as necessary, from asking to

[8] This might be mitigated by means of using TF-idf (Term frequency – Inverse document frequency) when accounting for the words used in the dictionary.

[9] We use an open-source library in Python, in particular *BertForSequenceClassification*. This library has a maximum size of the input text to analyzed, limited to 512 tokens. In Section 4 we will further comment on this limitation.

[10] As evidenced by Haque et al. (2022) the majority of the early adopters have expressed overwhelmingly positive sentiments on this model, stating that "this is the most disruptive technology" they have ever experienced with.

summarize a text, to make up a story, to translate a sentence or to calculate the sentiment. The flexibility in determining the task and ChatGPT's ability to understand it allows us to ask ChatGPT the sentiment in a text about specific topics (like CBDC) and from various angles if necessary. Also, we can interact with ChatGPT through the OpenAI API, which is very convenient because we can automate the process of asking ChatGPT to calculate sentiment on a certain topic across thousands of texts.

The selection of the prompt (usually known as prompt engineering) is very important, as different prompts could yield different results, depending as well on the fine-tuning of the hyper-parameters of the model.[11] In the prompt we will ask ChatGPT to compute the sentiment between -1 and 1, therefore we will not need any further transformation.

Our benchmark prompt for ChatGPT is as follows:

> "Compute the sentiment score towards central bank digital currencies, measured between -1 and 1, of a given text. The response should be just a float number, no text. The text is as follows: [...]"

Then, we ask an additional prompt which is:

> "Justify your answer"

We have performed robustness analysis with different prompts. At the time of writing, the only model version running behind ChatGPT was GPT-3.5 that has a limitation of 4,000 tokens per prompt.[12]

Here is an example of how ChatGPT works with the aforementioned prompt with two small texts from the Monetary Authority of Singapore.

Paragraph 1: *Third, a digital Singapore dollar could mitigate against the encroachment of privately issued stablecoins or foreign CBDCs in Singapore's payments landscape. As these global digital currencies enter our market and become widely accessible in the future, they could potentially displace the use of the Singapore dollar in domestic retail transactions. A digital Singapore dollar issued by MAS that is congruent with the needs of a digitalised economy could go some way to mitigate this risk.*

- ChatGPT: This is the score: 0.6
- ChatGPT (justify your answer): The text is discussing the possible benefits of a digital Singapore dollar issued by the MAS, and how it could help to prevent the displacement of the Singapore dollar in domestic retail transactions. The sentiment is positive towards the idea of a digital Singapore dollar

Paragraph 2: *Retail CBDCs can potentially pose significant risks to monetary and financial stability. There could be some disintermediation of the banks, particularly during stress periods if people can switch deposits into risk-free central bank money at the "click of a button". Even in normal times, if people held a significant portion of their deposits in the form of digital Singapore dollars with MAS, it would considerably reduce our banks' capacity to make loans. But we can likely manage these risks by designing the retail CBDC with sensible safeguards, such as stock and flow caps on the amount of digital Singapore dollars that anyone is allowed to place with MAS.*

- ChatGPT: This is the score: -0.6
- ChatGPT (justify your answer): The text discusses the potential risks of retail CBDCs, which could lead to disintermediation of banks and reduced capacity for loans. While the text does acknowledge that these risks can be managed, the overall sentiment is negative.

[11] We have used the *davinci-03* engine.

[12] OpenAI Technical Report states that GPT-4 will have a limitation of 32,000 tokens per prompt, but it was not available via API.

3.4. *Workflow*

As we said before, there are a total of 351 documents in our corpus. These documents are very heterogeneous, with different sizes. In we show the description of the size of these texts. There are texts with less than 1,000 words and others with more than 10,000. The average size is around 3,500 words, and almost half of the documents have more than 3,000 words. While we can apply the dictionary-based method to texts of any size, both BERT and ChatGPT have limitations, of 500 tokens and 4,000 tokens respectively.[13] And in the case of ChatGPT, both the prompt and answer tokens are included in the limitation. Therefore, keeping in mind the limitations of BERT and GPT, we have decided to perform our benchmark analysis splitting the documents into smaller pieces per document. This had an additional advantage because smaller texts are easy for a human to inspect, which allows us to immediately check how the dictionary-based and LLMs are performing.

We have decided to split each speech into paragraphs. Since the speeches cover many different topics other than CBDC (e.g.: COVID, financial stability, or macroeconomic forecasting), we select only relevant paragraphs where a number of keywords are mentioned. In particular, in line with Tian et al. (2022) the keywords are CBDC, central bank digital currency (currencies), central bank digital money, central-bank-issued digital money, digital euro, and central-bank-issued digital currency (currencies).[14] Once a relevant paragraph is selected, we calculate the sentiment with the three methods (Dictionary-based, BERT and ChatGPT). None of the paragraphs was bigger than 500 words (the average size of the selected paragraphs is 80 words), meaning that they could be analyzed by all techniques. In Table 2 we show the description of the size of the resulting text after applying the keywords, and the number of paragraphs, so that we can have an idea of the size of the texts fed into the models. Once all the paragraphs in the document have been analyzed, we calculate the document sentiment as the average of the paragraphs sentiments. We repeat this process for all the documents in our corpus. In Figure 1 we summarize the workflow.

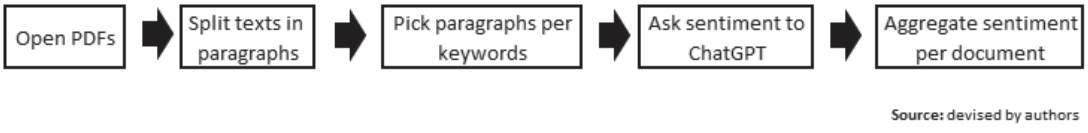

Figure 1. Workflow: procedure to automate the extraction of relevant text from the listed url of speeches provided in an Excel file in Auer et al. (2020). Step by step we show the methodology pursued to measure the sentiment of the corpus of text.

Table 2. Size of corpus: descriptive statistics of the corpus of texts, collected from Auer et al. (2020). First, the number of words in the full document, then the size of the relevant pieces of text (connected to CBDC), and finally, the number of paragraphs.

	Number of words, all text	Number of words, selected text	Number of paragraphs
Mean	3,513	558	7.11
Std	3,861	665	9.33
Min	529	21	1
0.25	1,894	123	1
0.5	2,895	278	3
0.75	3,872	699	9
Max	43,823	3,753	49

[13] A token is a single unit of text, like words, numbers, or punctuation marks, separated by white space or other delimiters

[14] We could perform this selection with ChatGPT as well. ChatGPT would be able to capture if the paragraph is talking about CBDC, but this will imply more queries and higher price through the API. A practical strategy could involve the use of semantic search, based on pre-trained embeddings, to select paragraphs most likely to address CBDCs. These selected paragraphs could then be fed into the API of ChatGPT.

4. Results

In this Section we analyze the sentiment on CBDC obtained with the three techniques: Dictionary-based, BERT and ChatGPT. First, in Section 5.1 we compare the results and the evolution over time, showing that sentiment on CBDC is clearly upward trending. Then, in Section 5.2 we compare the results obtained using NLP with the human expert analysis performed by BIS economists, showing that ChatGPT is closest to the labeled data. In Section 5.3, we perform different robust analyses such as different prompts, and increasing the size of the text. In the Appendix we show some more examples to support why ChatGPT can capture better the sentiment.

4.1. *Evolution of sentiment*

We have seen that all three techniques show a positive evolution of sentiment towards CBDC. But, which technique is more accurate? To assess the accuracy, we aim to calculate if they are comparable to what a human would do. To do this, we perform three exercises. We first calculate the correlation between the sentiment score of each technique with that reported by human experts, in particular the economists working at BIS, as stated in Auer et al. (2020). Here the authors used expert judgment to score each document with a value of -1, 0 or +1. Second, in the Appendix we look at some of the cases where there is major disagreement between the three techniques.

We can look into how sentiment evolves over time in several ways. First, in Figures 2 and 3, we group the documents by year (on the left) and by quarter (on the right), and take the average of the documents in that period. We must bear in mind that there is a different number of documents in each year and quarter (as we saw in Figure 1 above), so the figures may be biased by moments of time with a higher or lower number of documents. The annual and quarterly average sentiment shows a growing trend from 2017 onwards, for all techniques. Polarity seems to be more volatile and the fact that the sentiment of ChatGPT is greater than BERT, which in turn is greater than Polarity, is present in almost the entire sample. Sentiment for ChatGPT moves from 0.2 to 0.4, for BERT it moves around 0.1 and 0.2, and for Polarity it is much more volatile.

This analysis has the limitation that the number of documents in each quarter and year is different. For this reason, we carry out a moving average analysis, the results of which are shown in Figure 4. There, we show, document by document, the moving average of sentiment using 20 documents as a rolling window. The figure confirms that sentiment towards CBDC appears to be more positive as time goes by, based on all three techniques. For example, ChatGPT starts at about 0.2 and ends at 0.4, BERT starts at 0.1 and ends at 0.2, and Polarity is the most extreme case, starting from -0.1 to 0.3. Lower initial sentiment on CBDC is in line with the idea that central banks have started to talk cautiously about the systemic implications of CBDC (Barontini and Holden 2019). But over time, we see an increasing trend in sentiment, reflecting

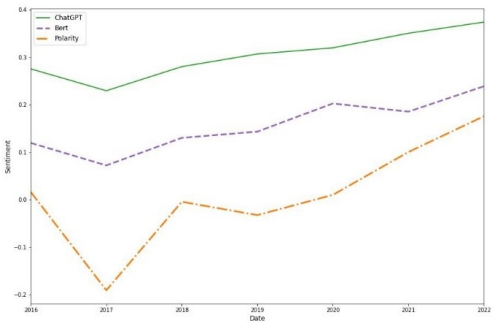

Figure 2. Sentiment. Avg year

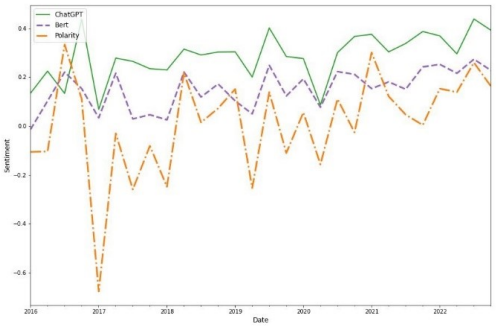

Figure 3. Sentiment. Avg quarter

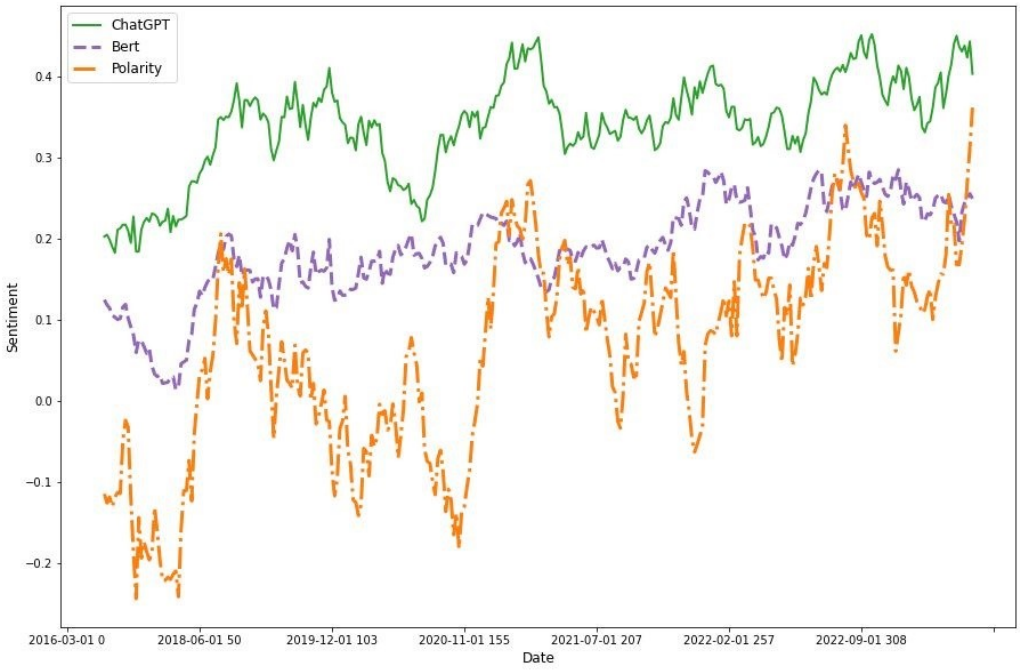

Figure 4. Sentiment. Moving Avg

that several central banks have become enthusiastic about exploring the idea of issuing a CBDC.

Table 3. Correlation with expert judgment. Benchmark exercise: correlation between the sentiment obtained by the three techniques with the sentiment labeled in Auer et al. (2020), for full texts, and for larger texts, combining paragraphs.

	All text	**+1 paragraph (242 docs)**
ChatGPT	0.34***	0.44***
BERT	0.29***	0.34***
Dictionary-based	0.21***	0.22***

4.2. *Comparing with humans and labeled data*

We have seen that all three techniques show a positive evolution of sentiment towards CBDC. But, which technique is more accurate? To assess the accuracy, we aim to calculate if they are comparable to what a human would do. To do this, we perform three exercises. We first calculate the correlation between the sentiment score of each technique with that reported by human experts, in particular the economists working at BIS, as stated in Auer et al. (2020). Here the authors used expert judgment to score each document with a value of -1, 0 or +1. Second, in the Appendix we look at some of the cases where there is major disagreement between the three techniques.[15]

In Table 3 we show the correlation between the sentiment obtained by the three techniques with the sentiment labeled in Auer et al. (2020), for full texts, and for larger texts, combining paragraphs. These correlations are in all cases positive and significant, which indicate that the techniques are capturing up to some level the sentiment

[15]We leave for further Research how to assess the predictability each sentiment score. E.g.: using a random forest model we could predict the sentiment score based on various characteristics, such as country of origin, time dummies, economic policy uncertainty index, interest in CBDC on Google web searches, and Bitcoin prices.

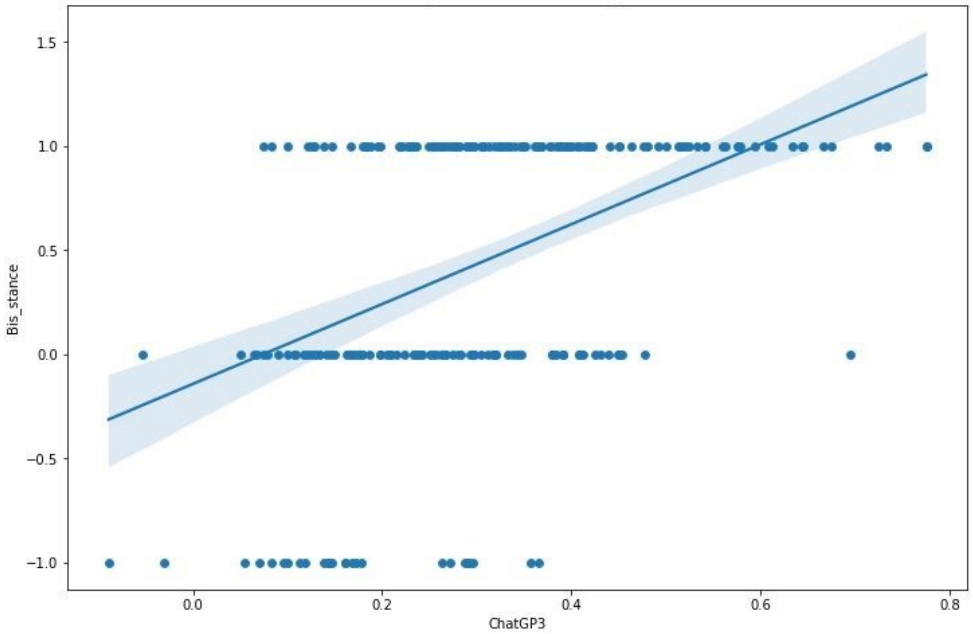

Figure 5. Correlation between ChatGPT and expert judgment: correlation between ChatGPT and human labels for texts with more than 100 words.

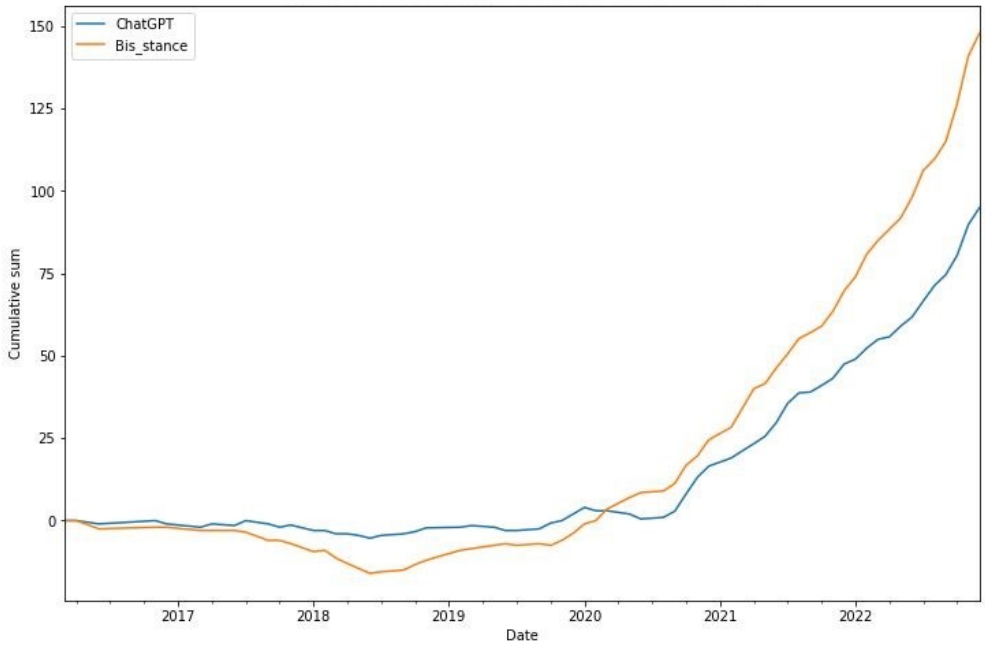

Figure 6. Cumulative sum ChatGPT vs expert judgment: comparing ChatGPT with human labels. The orange line is the BIS sentiment (Auer et al. 2020) and the blue line is ChatGPT.

better, at least compared to BERT and Polarity. Figure 5 shows the correlation between ChatGPT and human labels for texts with more than 100 words, where this relationship can be visually appreciated.

Finally, we compare the temporal evolution of the sentiment of ChatGPT with that of labeled data from human experts. To do this, we reconstructed the graph made in Auer et al. (2020), in which they plot the cumulative sum of sentiment month by month. To compare with ChatGPT, we translate the ChatGPT score to -1, 0, and 1, classifying the bottom 10% scores as -1, the top 50% scores as +1, and the rest of the scores as 0 (following the distribution -1, 0 and 1 of the expert judgment made by the human experts). The result is shown in Figure 6, with the orange line being for BIS

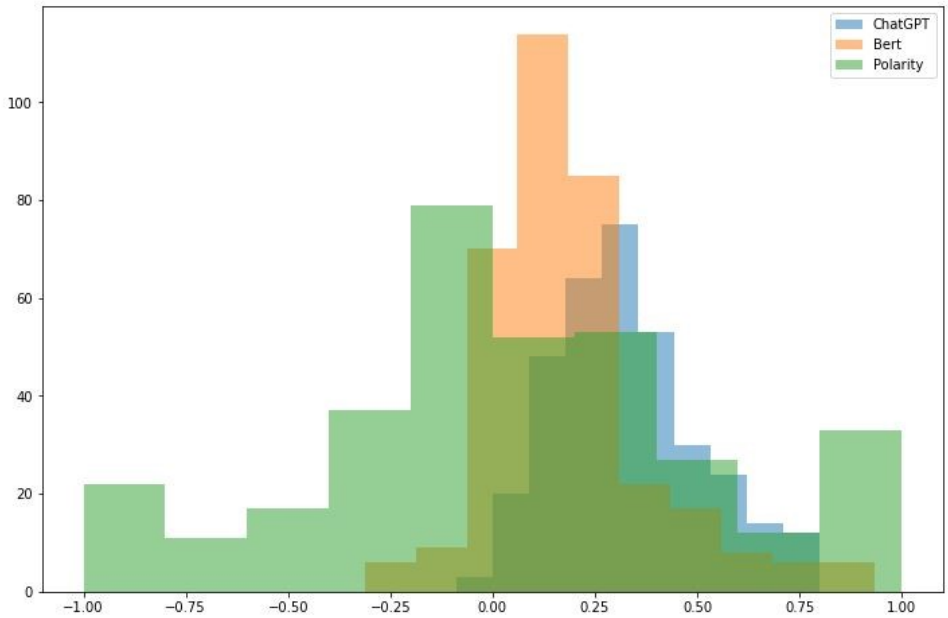

Figure 7. Histogram three techniques with different techniques: histogram of the sentiments per document provided by the three techniques

sentiment (Auer et al. 2020) and the blue line for ChatGPT. It can be seen that the trend changes are very similar. The similarity between the labeled data and ChatGPT is remarkable, especially considering that the prompt we used was quite generic, and it could be further tailored to capture BIS preferences.

Not only are the ChatGPT sentiment scores more similar to the expert labeled data, but it appears that the tone distribution of both ChatGPT and BERT are more human-like, while the polarity distribution differs. In Figure 7 we plot the histogram of the sentiments per document provided by the three techniques, and in Figure 8 we show the histogram of the typical scores provided by humans, as shown in Guo et al. (2023). The polarity histogram follows a nearly symmetrical distribution, whereas humans tend to follow a distribution closer to a Chi-square. Guo et al. (2023) conducted comprehensive human evaluations and linguistic analyses of the content generated by ChatGPT in comparison to that of humans, and it revealed interesting results. First, the authors find that the proportion of neutral emotions is higher for both humans and ChatGPT, which is in line with our results in this study. However, ChatGPT generally expresses less negative feelings than humans. This is also the case in our study. Therefore, a limitation of ChatGPT to analyze sentiment can occur when evaluating negative postures, since humans express more negative emotions than ChatGPT.

4.3. *Robustness analysis: Different prompts and larger texts*

Is ChatGPT capturing correctly the increasing sentiment of central banks towards CBDC, or it is by default more optimistic than BERT and Dictionary-based methods? Do our benchmark results hold for larger texts? In this Section we propose two analyses to test these questions.

4.3.1. Robustness analysis. Prompts for crypto

In order to test whether ChatGPT is more optimistic than BERT or Polarity by default, we perform a robustness analysis in which we repeat the workflow described in Figure 1 but select paragraphs according to the following keywords: crypto(s), crypto asset(s), crypto-asset(s), Bitcoin. We compute the sentiment over those paragraphs, and we use as a prompt for ChatGPT:

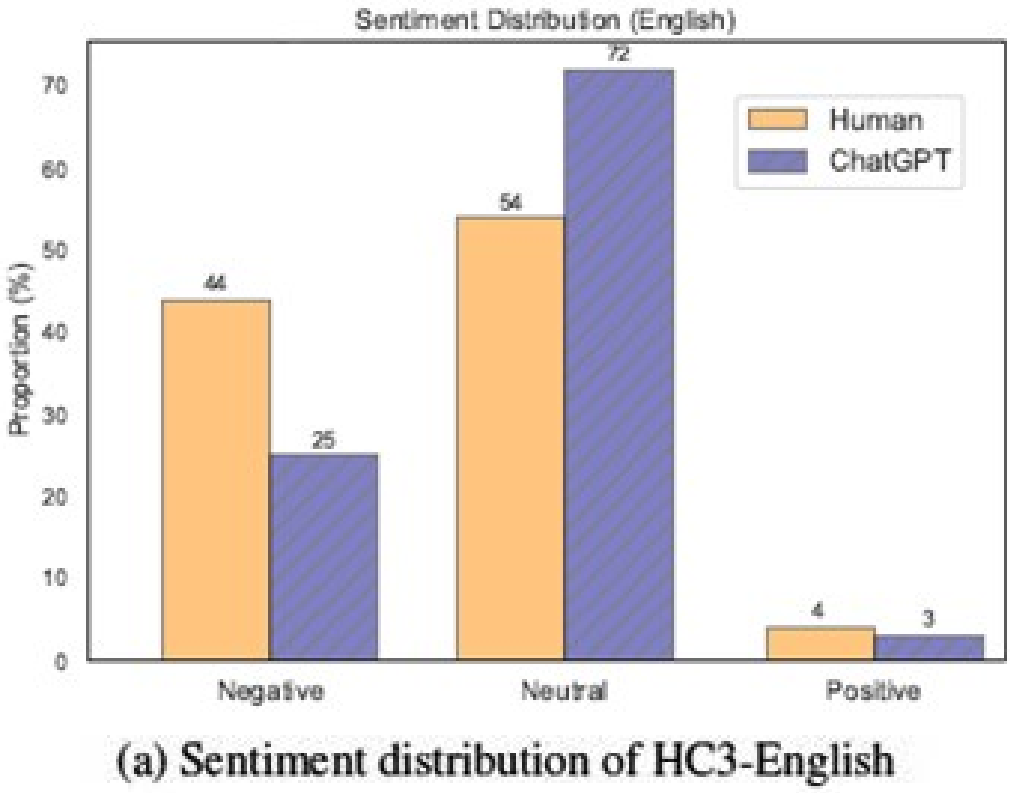

Figure 8. Guo et al. (2023): histogram of the typical scores provided by humans, as shown in Guo et al. (2023).

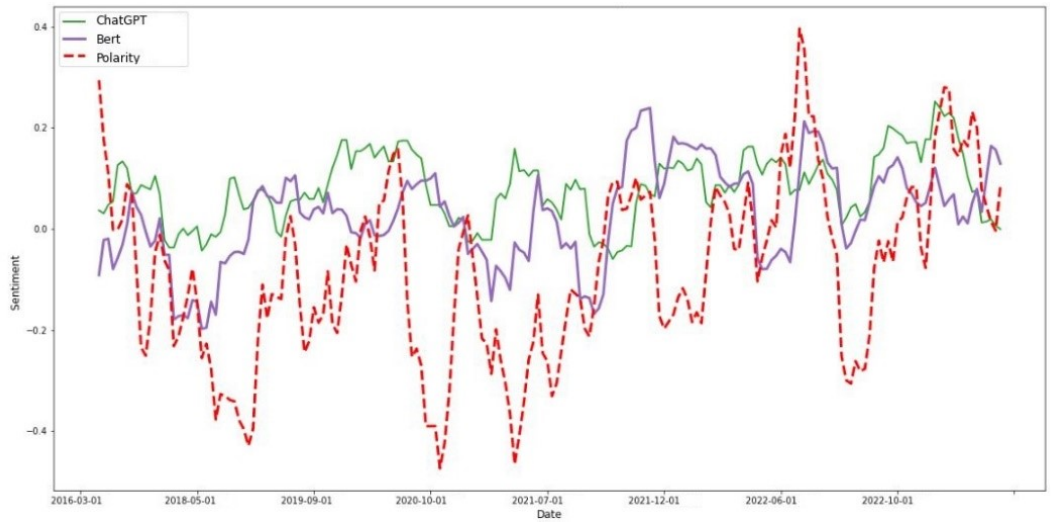

Figure 9. Sentiment scores towards crypto: results of the rolling window (moving average) analysis for the three NLP techniques, applied to crypto analysis.

"Compute the sentiment score towards crypto assets, measured between -1 and 1, of a given text. The response should be just a float number, no text. The text is as follows: [...]"

The results of the rolling window analysis for the three techniques can be seen in Figure 9. For crypto assets the sentiment does not display an upward trend behavior for any of the techniques, as expected.[16] Again, Polarity is the more volatile, and although on average is always below BERT and ChatGPT, there are moments in time in

[16]See for instance the list of speeches of the European Central Bank on crypto-assets.

which sentiment with Polarity is higher than the one from the LLMs. Moreover, ChatGPT (green) does not always indicate a higher sentiment that the other techniques. Therefore, the fact that ChatGPT captures more sentiment towards CBDC seems to be genuine, and is not just a result of the fact that ChatGPT could be less negative overall (Figure 8).

4.3.2. Robustness analysis. Larger texts

As we showed in Table 2, the average text size after selecting CBDC keywords is less than 300 words. Technically we could fit all texts with less than 500 words directly to BERT and ChatGPT without restrictions. In this robustness check, instead of evaluating the paragraphs one by one, we merge them (in texts with more than 1 paragraph) until we create big paragraphs of 500 words. The problem with this is that we might put together paragraphs that do not fit necessarily (they could belong to different parts of the texts). This does not matter for Polarity because it acts as a bag-of-words (the order of words does not matter) but might have implications for the LLMs. On the other hand, LLMs might benefit from having access to bigger texts. We show the correlation between labeled data and the three techniques with larger texts in Table 4. It can be seen that in larger texts Polarity does not improve its results, while BERT and ChatGPT improve it, specially ChatGPT. This is another indicator that ChatGPT is able to capture better the sentiment, and might benefit from the inclusion of larger texts. It is worth mentioning that GPT4 would be able to read texts of up to 32,000 tokens. However, at the time of writing, GPT4 was not accessible through the API.

Table 4. Correlation with expert judgment. Larger texts: correlation between labeled data and the three NLP techniques with larger texts.

	Texts of at least 500 words
ChatGPT	0.50 ***
Bert	0.38 ***
Dictionary	0.25 **

5. Conclusion and further work

Currently, central banks issue statements and opinions on topics that are not related to those traditionally associated with central bank communication (Hansson 2021). Recent studies suggest that these communications are important for financial markets, even when the topics discussed are not directly linked to monetary policy (Fortes and Le Guenedal 2020). One of these new topics is the discussion on the need and desirability of a new type of money: central bank digital currencies (CBDCs).

Our work investigates the sentiment of central banks' communication about the issuance of digital money through CBDC. For the first time in the literature we use LLMs (BERT and ChatGPT) to compute a sentiment score of these speeches and reports, and we compare them to traditional dictionary-based techniques and human labels. First, we find that the sentiment towards CBDC seems to be increasing from 2017 onwards according to both LLMs and dictionary-based techniques, but the latter are more volatile. Second, we compare the sentiment obtained by the three techniques with the one provided by experts, and we find that ChatGPT is closer to labeled data than BERT or Dictionary-based methods. If we select larger texts, the advantage of ChatGPT over the other techniques increase.

Therefore, we want to point out that this study provides central banks with insights on how central bank communications might be perceived by selected audiences. However, the immense size of the available LLMs brings two new risk factors. First,

interpretability of the results is a challenge currently being under scrutiny of regulators, and a field where developers are working, aiming to identify which parts of an LLM are responsible for which of its behaviors.[17] Additionally, the use of LLMs also raises concerns about third-party dependencies and the potential electrical and environmental cost of keeping these models online for everyone to access (Strubell et al. 2019).

Finally, we leave some indications for further research. While ChatGPT seems to better capture the sentiment towards CBDC than other NLP alternatives, we still need to assess the importance of prompt engineering when defining the task for ChatGPT, like changing its content, length, etc. In this article we have carried out several experiments in this direction, with prompts that exclude CBDC from the input statement (see Appendix section 7.3). On top of it, it could be worthwhile to extend our robustness analysis on the use of LLM techniques to compute the sentiment on other contexts, such as we did with crypto-assets (Section 5.3.1.). This way we could understand if ChatGPT is inherently more positive than BERT, or if dictionary-based methods are always more volatile. Also, future research could extend the analysis to other LLM techniques, like GPT4, XLNet, LLaMA, or T5. Finally, since our analysis provides a continuous measure of sentiment values, document by document, we could study which are the determinants of this sentiment index, and in particular, if there are factors that can explain why it fluctuates and why it increases. Are these sentiment values affected by changes in the crypto market? Or by the appearance of private digital currency initiatives like Libra? These analyses could help policymakers and market participants to better assess the likelihood of CBDC issuance, which would facilitate transparency for citizens about the current state of this complex debate.

In short, the use of these tools seems to have great potential in the field of sentiment analysis in general, although we must weigh in the risks and analyze the advantages and limitations of its usage case by case.

[17] See for instance the guidelines from OpenAI.

References

Adrian, Tobias, and Tommaso Mancini-Griffoli. (2021). "The rise of digital money". Annual *Review of Financial Economics*, 13(1), pp. 57-77. https://doi.org/10.1146/annurev-financial-101620-063859

Albrizio, Silvia, Juan Carlos Berganza e Iván Kataryniuk. (2017). "El seguro de desempleo federal en Estados Unidos". *Boletín Económico - Banco de España*, 2/2017, Artículos Analíticos. https://repositorio.bde.es/handle/123456789/8263

Andolfatto, David. (2020). "Assessing the Impact of Central Bank Digital Currency on Private Banks". *The Economic Journal*, 131(634), pp. 525-540. https://doi.org/10.1093/ej/ueaa073

Ash, Elliott, and Stephen Hansen. (2023). "Text algorithms in economics". Unpublished manuscript.

Auer, Raphael, Holti Banka, Nana Yaa Boakye-Adjei, Ahmed Faragallah, Jon Frost, Harish Natarajan and Jermy Prenio. (2022a). "Central bank digital currencies: a new tool in the financial inclusion toolkit?". FSI Insights on Policy Implementation, 41, Bank for International Settlements. https://www.bis.org/fsi/publ/insights41.pdf

Auer, Raphael, and Rainer Boehme. (2020). "The technology of retail central bank digital currency". *BIS Quarterly Review*. https://EconPapers.repec.org/RePEc:bis:bisqtr:2003j

Auer, Raphael, Giulio Cornelli and Jon Frost. (2020). "Rise of the central bank digital currencies: drivers, approaches and technologies". BIS Working Papers, 880, Bank for International Settlements. https://www.bis.org/publ/work880.pdf

Auer, Raphael, Jon Frost, Leonardo Gambacorta, Cyril Monnet, Tara Rice and Hyun Song Shin. (2022b). "Central bank digital currencies: Motives, economic implications, and the research frontier". *Annual Review of Economics*, 14(1), pp. 697-721. https://doi.org/10.1146/annurev-economics-051420-020324

Barontini, Christian, and Henry Holden. (2019). "Proceeding with caution - a survey on central bank digital currency". BIS Papers, 101, Bank for International Settlements. https://www.bis.org/publ/bppdf/bispap101.pdf

Barrdear, John, and Michael Kumhof. (2016). "The Macroeconomics of Central Bank Issued Digital Currencies". Staff Working Papers, 605, Bank of England. https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2016/the-macroeconomics-of-central-bank-issued-digital-currencies.pdf

Bholat, David, Nida Broughton, Janna Ter Meer and Eryk Walczak. (2019). "Enhancing central bank communications using simple and relatable information". *Journal of Monetary Economics*, 108(C), pp. 1-15. https://EconPapers.repec.org/RePEc:eee:moneco:v:108:y:2019:i:c:p:1-15

Blinder, Alan S. (2018). "Through a crystal ball darkly: The future of monetary policy communication". *AEA Papers and Proceedings*, 108, pp. 567-571. https://www.aeaweb.org/articles?id=10.1257/pandp.20181080

Blinder, Alan S., Michael Ehrmann, Marcel Fratzscher, Jakob De Haan and David-Jan Jansen. (2008). "Central bank communication and monetary policy: A survey of theory and evidence". *Journal of Economic Literature*, 46(4), pp. 910-945. https://doi.org/10.1257/jel.46.4.910

Born, Benjamin, Michael Ehrmann and Marcel Fratzscher. (2014). "Central bank communication on financial stability". *Economic Journal*, 124(577), pp. 701-734. https://EconPapers.repec.org/RePEc:wly:econjl:v:124:y:2014:i:577:p:701-734

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro and Yi Zhang. (2023). "Sparks of Artificial General Intelligence: Early experiments with GPT-4". https://arxiv.org/abs/2303.12712

Burlon, Lorenzo, Carlos Montes-Galdón, Manuel A. Muñoz and Frank Smets. (2022). "The optimal quantity of CBDC in a bank-based economy". Working Paper Series, 2689, European Central Bank. https://ideas.repec.org/p/ecb/ecbwps/20222689.html

Campbell, Jeffrey R., Filippo Ferroni, Jonas D. M. Fisher and Leonardo Melosi. (2019). "The limits of forward guidance". *Journal of Monetary Economics,* 108(C), pp. 118-134. https://doi.org/10.1016/j.jmoneco.2019.08.009

Cole, Stephen J. (2021). "Learning and the Effectiveness of Central Bank Forward Guidance". *Journal of Money, Credit and Banking,* 53(1), pp. 157-200. https://doi.org/10.1111/jmcb.12696

Conlon, Thomas, Shaen Corbet, Greg Hou, Yang Hu, Charles James Larkin and Les Oxley. (2022). "To CBDC or Not to CBDC. That Is the Question." August 31, 2022. http://dx.doi.org/10.2139/ssrn.4205402

Davoodalhosseini, Seyed Mohammadreza. (2022). "Central bank digital currency and monetary policy". *Journal of Economic Dynamics and Control,* 142(C). https://doi.org/10.1016/j.jedc.2021.104150

Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". https://arxiv.org/abs/1810.04805

Dowling, Michael, and Brian Lucey. (2023). "ChatGPT for (Finance) research: The Bananarama Conjecture". *Finance Research Letters,* 53(C). https://doi.org/10.1016/j.frl.2023.103662

Fernández-Villaverde, Jesús, Daniel Sanches, Linda Schilling and Harald Uhlig. (2021). "Central Bank Digital Currency: Central Banking For All?". *Review of Economic Dynamics,* 41, pp. 225-242. https://doi.org/10.1016/j.red.2020.12.004

Ferrari Minesso, Massimo, Arnaud Mehl and Livio Stracca. (2022). "Central bank digital currency in an open economy". *Journal of Monetary Economics,* 127(C), pp. 54-68. https://doi.org/10.1016/j.jmoneco.2022.02.001

Fortes, Roberta, and Theo Le Guenedal. (2020). "Tracking ECB's communication: Perspectives and implications for financial markets". MPRA paper, University Library of Munich, Germany. https://EconPapers.repec.org/RePEc:pra:mprapa:108746

Gentzkow, Matthew, Bryan Kelly and Matt Taddy. (2019). "Text as data". *Journal of Economic Literature,* 57(3), pp. 535-74. https://doi.org/10.1257/jel.20181020

Goldsmith-Pinkham, Paul, Beverly Hirtle and David O. Lucca. (2016). "Parsing the content of bank supervision". Staff Reports, 770, Federal Reserve Bank of New York. https://ideas.repec.org/p/fip/fednsr/770.html

Gorjón Rivas, Sergio. (2022). "Mercados financieros mayoristas y divisas digitales: avanzando en la tokeneización del dinero de banco central". *Revista de Estabilidad Financiera - Banco de España,* 42, pp. 91-105. https://repositorio.bde.es/handle/123456789/21560

Guo, Biyang, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue and Yupeng Wu. (2023). "How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection". https://arxiv.org/abs/2301.07597

Guthrie, Graeme, and Julian Wright. (2000). "Open mouth operations". *Journal of Monetary Economics,* 46(2), pp. 489-516. https://EconPapers.repec.org/RePEc:eee:moneco:v:46:y:2000:i:2:p:489-516

Haldane, Andrew, and Michael McMahon. (2018). "Central bank communications and the general public". *AEA Papers and Proceedings,* 108, pp. 578-583. https://doi.org/10.1257/pandp.20181082

Hansen, Anne Lundgaard, and Sophia Kazinnik. (2023). "Can chatGPT decipher fedspeak?". http://dx.doi.org/10.2139/ssrn.4399406

Hansen, Stephen, Michael McMahon and Andrea Prat. (2018). "Transparency and deliberation within the fomc: A computational linguistics approach". *The Quarterly Journal of Economics*, 133(2), pp. 801-870. https://EconPapers.repec.org/RePEc:oup:qjecon:v:133:y:2018:i:2:p:801-870

Hansen, Stephen, Michael McMahon and Matthew Tong. (2019). "The long-run information effect of central bank communication". *Journal of Monetary Economics*, 108(C), pp. 185-202. https://doi.org/10.1016/j.jmoneco.2019.09.002

Hansson, Magnus. (2021). "Evolution of topics in central bank speech communication". https://arxiv.org/abs/2109.10058

Haque, Mubin Ul, Isuru Dharmadasa, Zarrin Tasnim Sworna, Roshan Namal Rajapakse and Hussain Ahmad. (2022). "'I think this is the most disruptive technology': Exploring Sentiments of ChatGPT Early Adopters using Twitter Data". https://arxiv.org/abs/2212.05856

Hayo, Bernd, and Johannes Zahner. (2023). "What is that noise? Analysing sentiment-based variation in central bank communication". *Economics Letters*, 222(C). https://doi.org/10.1016/j.econlet.2022.110962

Kiff, John, Jihad Alwazir, Sonja Davidovic, Aquiles Farias, Ashraf Khan, Tanai Khiaonarong, Majid Malaika, Hunter Monroe, Nobu Sugimoto, Hervé Tourpe and Peter Zhou (2020). "A survey of research on retail central bank digital currency". IMF Working Papers, 2020/104, International Monetary Fund. http://dx.doi.org/10.2139/ssrn.3639760

Li, Jiaqi. (2023). "Predicting the demand for central bank digital currency: A structural analysis with survey data". *Journal of Monetary Economics*, 134(C), pp. 73-85. https://doi.org/10.1016/j.jmoneco.2022.11.007

Loughran, Tim, and Bill McDonald. (2011). "When is a liability not a liability? Textual Analysis, Dictionaries, and 10-Ks". *Journal of Finance*, 66(1), pp. 35-65. https://EconPapers.repec.org/RePEc:bla:jfinan:v:66:y:2011:i:1:p:35-65

McKay, Alisdair, Emi Nakamura and Jón Steinsson. (2016). "The power of forward guidance revisited". *American Economic Review*, 106(10), pp. 3133-3158. https://doi.org/10.1257/aer.20150063

Romero Ugarte, José Luis, Abel Sánchez Martín, Carlos Martín Rodríguez and Justo Arenillas Cristóbal. (2021). "Implicaciones de una moneda digital soberana mayorista apoyada en tecnología de registros distribuidos para las infraestructuras del mercado financiero". *Revista de Estabilidad Financiera - Banco de España*, 40, pp. 161-178. https://repositorio.bde.es/handle/123456789/16737

Scharnowski, Stefan. (2022). "Central bank speeches and digital currency competition". *Finance Research Letters*, 49(C). https://doi.org/10.1016/j.frl.2022.103072

Siklos, Pierre, Samantha St Amand and Joanna Wajda. (2018). "The Evolving Scope and Content of Central Bank Speeches". CIGI Papers, 202, Centre for International Governance Innovation. https://www.cigionline.org/static/documents/documents/Paper%20No.202web_0.pdf

Strubell, Emma, Ananya Ganesh and Andrew McCallum. (2019). "Energy and Policy Considerations for Deep Learning in NLP". In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics, pp. 3645-3650. https://doi.org/10.18653/v1/P19-1355

Tarlin, Solomon H. (2021). "The Future of Cash". Discussion Paper, 21-03, Federal Reserve Bank of Philadelphia. https://doi.org/10.21799/frbp.dp.2021.03

Tian, Shu, Bo Zhao and Resi Ong Olivares. (2023). "Cybersecurity risks and central banks' sentiment on central bank digital currency: Evidence from global cyberattacks". *Finance Research Letters*, 53(C). https://doi.org/10.1016/j.frl.2022.103609

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin. (2017). "Attention is all you need". In Guyon, I., U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Wang, Yizhi, Brian M. Lucey, Samuel A. Vigne and Larisa Yarovaya. (2022). "The Effects of Central Bank Digital Currencies News on Financial Markets". *Technological Forecasting and Social Change*, 180(C). https://doi.org/10.1016/j.techfore.2022.121715

Yang, Yi, Mark Christopher Siy UY and Allen Huang. (2020). "FinBERT: A Pretrained Language Model for Financial Communications". https://arxiv.org/abs/2006.08097

Wang, Yizhi, Brian M. Lucey, Samuel A. Vigne and Larisa Yarovaya. (2022). "The Effects of Central Bank Digital Currencies News on Financial Markets". *Technological Forecasting and Social Change*, 180(C). https://doi.org/10.1016/j.techfore.2022.1

Yang, Yi, Mark Christopher Siy UY and Allen Huang. (2020). "FinBERT: A Pretrained Language Model for Financial Communications". https://arxiv.org/abs/2006.08097

6. Appendix

6.1. *Cross correlation between ChatGPT, BERT, and Dictionary*

In this section we are going to explore how the results of ChatGPT are similar to those of BERT and the dictionary. In Section 5.1 we showed that they all show an upward trend in sentiment about CBDC. In Table 5 we show the cross correlation between them. It can be seen that, as expected, BERT and ChatGPT have closer correlation structure. The one that looks the most like the dictionary is BERT. All these correlations are statistically significant. In Figure 10 we plot the correlation between ChatGPT and BERT. Each dot represents a document.

Table 5. Correlation sentiment between the three techniques: cross correlation between all three NLP techniques.

	ChatGPT	BERT	Dictionary-based
ChatGPT	**1**	**x**	**x**
BERT	**0.52*****	**1**	**x**
Dictionary-based	**0.29*****	**0.32*****	**1**

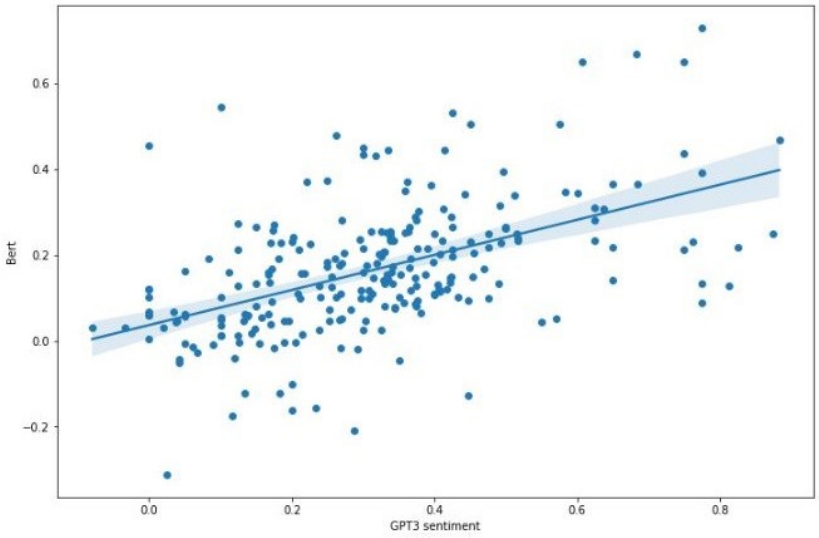

Figure 10. Correlation between ChatGPT and BERT: correlation between ChatGPT and BERT. Each dot represents a document.

It can be clearly seen that higher (lower) ChatGPT scores are related to higher (lower) BERT scores. Even so, we can see that the main source of disagreement between BERT and ChatGPT are several observations in which BERT catalogs the stance as neutral, but ChatGPT captures it as positive. That was expected. After all, as mentioned above, ChatGPT tends to be more positive than BERT when we ask about CBDC. Who is more right in these disagreements? These disagreements are due to the fact that ChatGPT can grasp subtler behaviors. In the next section we look at some examples.

6.2. *Examples of disagreements*

In this section we show three examples. The first two are ha a clearly positive tone towards CBDC, but only ChatGPT captures it correctly, BERT mistakenly stays neutral, and dictionary marks negative. In Example 1 there are clear indications that the tone is positive towards CBDC. The author says that "the aim would be to provide a kind of digital banknote, which would combine the security of banknotes issued by the central bank with the ease of use and innovation of dematerialised means of payment", and says that they are working on it. This suggests that the tone is positive, but BERT

marks it as neutral, the dictionary-based method (polarity) marks it as negative, which is clearly wrong. The same dynamics can be seen in Example 2.

Example 1: François Villeroy de Galhau: New Year wishes to the Paris financial Centre (2022-01-01)

"A forward looking euro means new banknotes after 2024, with a new design, and even more secure there are now only 17 counterfeits per million banknotes, making the euro one of the most secure currencies in the world. It is also a digital euro. The investigation phase for a "digital retail euro" was launched on 1 October last year, for a period of two years. The aim would be to provide a kind of "digital banknote", which would combine the security of banknotes issued by the central bank with the ease of use and innovation of dematerialised means of payment. The ECB and the Banque de France are working on all the associated challenges, and we have already conducted 9 experiments in Paris on an "interbank" digital currency. They will be continued in 2022, focusing on improving cross border payments".

- ChatGPT (justify your answer): 0.8 The text discusses the possibility of a digital euro, and the overall tone is positive. The text mentions the benefits of a digital euro, such as increased security and ease of use, and the ECB is conducting experiments to improve the system. There is no mention of negative consequences of a digital euro, so the sentiment score is positive
- BERT: This is the score: 0.07
- Polarity: This is the score: -0.44

Example 2: Mario Marcel: Chile's September 2020 Monetary Policy report (2020-10-01)

"Meanwhile, we have been closely monitoring all the developments associated with digital payment methods. As part of this follow-up, our participation in research groups of multilateral institutions has been active, especially on digital money issued by central banks. The digital issuance of central bank money is gaining more space in the discussion agenda at different international organizations, and the repercussions of the current sanitary crisis are accelerating this debate even more. Today, there are several serious projects being explored by central banks and international organizations on the subject. We believe that, in spite of both global and idiosyncratic contingencies, it is opportune to deepen our understanding and exploration of this type of instruments and related technologies".

- ChatGPT (justify your answer): 0.7. The sentiment score towards central bank digital currencies in this text is positive. The author is discussing the developments associated with digital payment methods and the central bank's participation in research groups exploring digital money. The author believes that central bank digital currencies are a serious topic worthy of exploration. This indicates a positive sentiment towards central bank digital currencies.
- BERT: This is the score: 0.1
- Polarity: This is the score: -0.5

Example 3: Burkhard Balz (Germany): Current challenges in an uncertain world (2022-03-01)

"Central banks, too, are maintaining intensive worldwide collaboration. We central bankers are facing similar challenges, be they finding the right monetary policy responses to developments in the real economy and financial sector or the introduction of a central bank digital currency – a topic that has been greatly occupying me."

- ChatGPT (justify your answer): 0.8. The sentiment score of this text is positive, as the author is discussing the various challenges that central bankers are facing and how they are working together to overcome them. The author specifically mentions the introduction of a central bank digital currency, which is an issue that has been occupying him, as a positive development. This shows that the

author is in favor of central bank digital currencies and has a positive sentiment towards them.

- BERT: This is the score: 0.0
- Polarity: This is the score: 0.3

6.3. *ChatGPT without asking specifically about CBDC*

Finally, we tested what happens if we ask ChatGPT the sentiment for phrases containing CBDC, without mentioning anything about CBDC in the prompt. The prompt would just be: *"What is the sentiment of the next sentence?"*. In this way we can better compare with BERT, since BERT does not allow to ask the feeling about specific concepts. In Figure Figure 11 we show the result. In green we show the original ChatGPT result (asking for CBDC), in orange and dotted we show the ChatGPT result without asking for CBDC, and in purple the original result from BERT (same as in Section 5). We can see that BERT and ChatGPT without CBDC are more similar to each other. This result supports that the greater positive effect found by ChatGPT when asked about CBDC is justified.

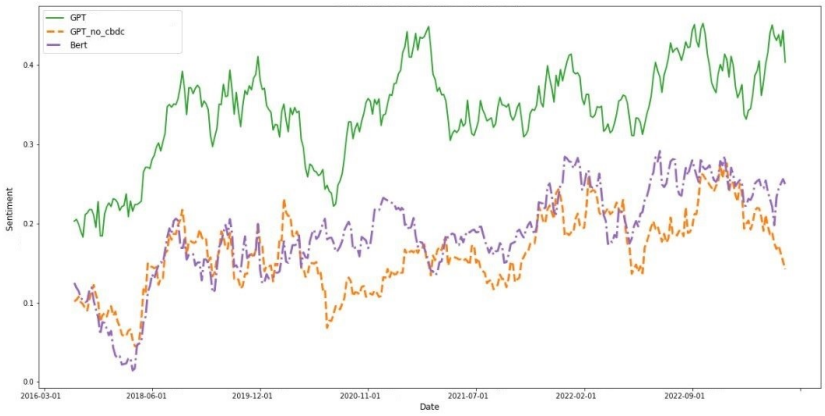

Figure 11. Sentiment. Moving average. (ChatGPT without CBDC reference): sentiment score of ChatGPT and BERT without mentioning CBDC in the prompt.

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

2215 JOSÉ MANUEL CARBÓ and SERGIO GORJÓN: Application of machine learning models and interpretability techniques to identify the determinants of the price of bitcoin.
2216 LUIS GUIROLA and MARÍA SÁNCHEZ-DOMÍNGUEZ: Childcare constraints on immigrant integration.
2217 ADRIÁN CARRO, MARC HINTERSCHWEIGER, ARZU ULUC and J. DOYNE FARMER: Heterogeneous effects and spillovers of macroprudential policy in an agent-based model of the UK housing market.
2218 STÉPHANE DUPRAZ, HERVÉ LE BIHAN and JULIEN MATHERON: Make-up strategies with finite planning horizons but forward-looking asset prices.
2219 LAURA ÁLVAREZ, MIGUEL GARCÍA-POSADA and SERGIO MAYORDOMO: Distressed firms, zombie firms and zombie lending: a taxonomy.
2220 BLANCA JIMÉNEZ-GARCÍA and JULIO RODRÍGUEZ: A quantification of the evolution of bilateral trade flows once bilateral RTAs are implemented.
2221 SALOMÓN GARCÍA: Mortgage securitization and information frictions in general equilibrium.
2222 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Accuracy of explanations of machine learning models for credit decisions.
2223 JAMES COSTAIN, GALO NUÑO and CARLOS THOMAS: The term structure of interest rates in a heterogeneous monetary union.
2224 ANTOINE BERTHEAU, EDOARDO MARIA ACABBI, CRISTINA BARCELÓ, ANDREAS GULYAS, STEFANO LOMBARDI and RAFFAELE SAGGIO: The Unequal Consequences of Job Loss across Countries.
2225 ERWAN GAUTIER, CRISTINA CONFLITTI, RIEMER P. FABER, BRIAN FABO, LUDMILA FADEJEVA, VALENTIN JOUVANCEAU, JAN-OLIVER MENZ, TERESA MESSNER, PAVLOS PETROULAS, PAU ROLDAN-BLANCO, FABIO RUMLER, SERGIO SANTORO, ELISABETH WIELAND and HÉLÈNE ZIMMER. New facts on consumer price rigidity in the euro area.
2226 MARIO BAJO and EMILIO RODRÍGUEZ: Integrating the carbon footprint into the construction of corporate bond portfolios.
2227 FEDERICO CARRIL-CACCIA, JORDI PANIAGUA and MARTA SUÁREZ-VARELA: Forced migration and food crises.
2228 CARLOS MORENO PÉREZ and MARCO MINOZZO: Natural Language Processing and Financial Markets: Semi-supervised Modelling of Coronavirus and Economic News.
2229 CARLOS MORENO PÉREZ and MARCO MINOZZO: Monetary Policy Uncertainty in Mexico: An Unsupervised Approach.
2230 ADRIAN CARRO: Could Spain be less different? Exploring the effects of macroprudential policy on the house price cycle.
2231 DANIEL SANTABÁRBARA and MARTA SUÁREZ-VARELA: Carbon pricing and inflation volatility.
2232 MARINA DIAKONOVA, LUIS MOLINA, HANNES MUELLER, JAVIER J. PÉREZ and CRISTOPHER RAUH: The information content of conflict, social unrest and policy uncertainty measures for macroeconomic forecasting.
2233 JULIAN DI GIOVANNI, MANUEL GARCÍA-SANTANA, PRIIT JEENAS, ENRIQUE MORAL-BENITO and JOSEP PIJOAN-MAS: Government Procurement and Access to Credit: Firm Dynamics and Aggregate Implications.
2234 PETER PAZ: Bank capitalization heterogeneity and monetary policy.
2235 ERIK ANDRES-ESCAYOLA, CORINNA GHIRELLI, LUIS MOLINA, JAVIER J. PÉREZ and ELENA VIDAL: Using newspapers for textual indicators: which and how many?
2236 MARÍA ALEJANDRA AMADO: Macroprudential FX regulations: sacrificing small firms for stability?
2237 LUIS GUIROLA and GONZALO RIVERO: Polarization contaminates the link with partisan and independent institutions: evidence from 138 cabinet shifts.
2238 MIGUEL DURO, GERMÁN LÓPEZ-ESPINOSA, SERGIO MAYORDOMO, GAIZKA ORMAZABAL and MARÍA RODRÍGUEZ-MORENO: Enforcing mandatory reporting on private firms: the role of banks.
2239 LUIS J. ÁLVAREZ and FLORENS ODENDAHL: Data outliers and Bayesian VARs in the Euro Area.
2240 CARLOS MORENO PÉREZ and MARCO MINOZZO: "Making text talk": The minutes of the Central Bank of Brazil and the real economy.
2241 JULIO GÁLVEZ and GONZALO PAZ-PARDO: Richer earnings dynamics, consumption and portfolio choice over the life cycle.
2242 MARINA DIAKONOVA, CORINNA GHIRELLI, LUIS MOLINA and JAVIER J. PÉREZ: The economic impact of conflict-related and policy uncertainty shocks: the case of Russia.
2243 CARMEN BROTO, LUIS FERNÁNDEZ LAFUERZA and MARIYA MELNYCHUK: Do buffer requirements for European systemically important banks make them less systemic?
2244 GERGELY GANICS and MARÍA RODRÍGUEZ-MORENO: A house price-at-risk model to monitor the downside risk for the Spanish housing market.

2245 JOSÉ E. GUTIÉRREZ and LUIS FERNÁNDEZ LAFUERZA: Credit line runs and bank risk management: evidence from the disclosure of stress test results.

2301 MARÍA BRU MUÑOZ: The forgotten lender: the role of multilateral lenders in sovereign debt and default.

2302 SILVIA ALBRIZIO, BEATRIZ GONZÁLEZ and DMITRY KHAMETSHIN: A tale of two margins: monetary policy and capital misallocation.

2303 JUAN EQUIZA, RICARDO GIMENO, ANTONIO MORENO and CARLOS THOMAS: Evaluating central bank asset purchases in a term structure model with a forward-looking supply factor.

2304 PABLO BURRIEL, IVÁN KATARYNIUK, CARLOS MORENO PÉREZ and FRANCESCA VIANI: New supply bottlenecks index based on newspaper data.

2305 ALEJANDRO FERNÁNDEZ-CEREZO, ENRIQUE MORAL-BENITO and JAVIER QUINTANA: A production network model for the Spanish economy with an application to the impact of NGEU funds.

2306 MONICA MARTINEZ-BRAVO and CARLOS SANZ: Trust and accountability in times of pandemic.

2307 NATALIA FABRA, EDUARDO GUTIÉRREZ, AITOR LACUESTA and ROBERTO RAMOS: Do Renewables Create Local Jobs?

2308 ISABEL ARGIMÓN and IRENE ROIBÁS: Debt overhang, credit demand and financial conditions.

2309 JOSÉ-ELÍAS GALLEGOS: Inflation persistence, noisy information and the Phillips curve.

2310 ANDRÉS ALONSO-ROBISCO, JOSÉ MANUEL CARBÓ and JOSÉ MANUEL MARQUÉS: Machine Learning methods in climate finance: a systematic review.

2311 ALESSANDRO PERI, OMAR RACHEDI and IACOPO VAROTTO: The public investment multiplier in a production network.

2312 JUAN S. MORA-SANGUINETTI, JAVIER QUINTANA, ISABEL SOLER and ROK SPRUK: Sector-level economic effects of regulatory complexity: evidence from Spain.

2313 CORINNA GHIRELLI, ENKELEJDA HAVARI, ELENA MERONI and STEFANO VERZILLO: The long-term causal effects of winning an ERC grant.

2314 ALFREDO GARCÍA-HIERNAUX, MARÍA T. GONZÁLEZ-PÉREZ and DAVID E. GUERRERO: How to measure inflation volatility. A note.

2315 NICOLÁS ABBATE, INÉS BERNIELL, JOAQUÍN COLEFF, LUIS LAGUINGE, MARGARITA MACHELETT, MARIANA MARCHIONNI, JULIÁN PEDRAZZI and MARÍA FLORENCIA PINTO: Discrimination against gay and transgender people in Latin America: a correspondence study in the rental housing market.

2316 SALOMÓN GARCÍA: The amplification effects of adverse selection in mortgage credit suply.

2317 METTE EJRNÆS, ESTEBAN GARCÍA-MIRALLES, METTE GØRTZ and PETTER LUNDBORG: When death was postponed: the effect of HIV medication on work, savings and marriage.

2318 GABRIEL JIMÉNEZ, LUC LAEVEN, DAVID MARTÍNEZ-MIERA and JOSÉ-LUIS PEYDRÓ: Public guarantees and private banks' incentives: evidence from the COVID-19 crisis.

2319 HERVÉ LE BIHAN, DANILO LEIVA-LEÓN and MATÍAS PACCE: Underlying inflation and asymmetric risks.

2320 JUAN S. MORA-SANGUINETTI, LAURA HOSPIDO and ANDRÉS ATIENZA-MAESO: Los números de la regulación sobre igualdad. Cuantificación de la actividad normativa sobre no discriminación en España y su relación con las brechas de género en el mercado de trabajo.

2321 ANDRES ALONSO-ROBISCO and JOSÉ MANUEL CARBÓ: Analysis of CBDC Narrative of Central Banks using Large Language Models.