

Analysis of CBDC Narrative by Central Banks using Large Language Models*



By Andrés Alonso Robisco and José Manuel Carbó
Banco de España

Keywords: ChatGPT, BERT, CBDC, digital money.

JEL codes: G15, G41, E58, C88.

Central banks are increasingly using verbal communication for policy making, 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 the attention from the international community. To better understand central banks' stance towards CBDCs, we used different natural language processing techniques on a set of central bank speeches. We found that the sentiment calculated by Large Language Models, and in particular by ChatGPT, is the one that most resembles the sentiment identified by human experts in those same speeches. Our study suggests that LLMs are an effective tool for improving sentiment measurements on specific policy texts, although they are not infallible and may be subject to new risks.

*This Policy Brief is based on [Banco de España Documentos de Trabajo. N.º 2321](#). 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.

1. Central Bank Digital Currency and sentiment

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 publish economic projections.

A topic that is gaining momentum within central banks communication is Central Bank Digital Currency (CBDC), a new type 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 deposits, wholesale deposits, or both. Depending on its final design, a CBDC could serve various purposes, such as payment system integrity, promoting financial inclusion, and fostering innovation, among others. But 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 et al 2022).

Many central banks are exploring the potential value of CBDC, though at various paces. Some are simply incorporating this topic into speeches, while a few are conducting pilots and live experimentation (Auer et al 2020). There are several approaches to the design of CBDCs, and these vary significantly between jurisdictions. Some possible design options are the choice of the central bank's operational function, the choice of infrastructure (DLT or centralized database), the access method (account-based or token-based), or cross-border payment links.

Given the uncertain design and function of CBDCs, analyzing central bank sentiment towards CBDC through speeches could offer insight into potential policy direction, improving transparency and market expectations. With this in mind, we have calculated the sentiment that central banks express towards CBDCs in a series of speeches, using natural language processing (NLP) techniques, specifically two large language models (LLM), such as ChatGPT and BERT, and traditional dictionary-based methods.

2. Dataset and methodology

Dataset

We have used a collection of central bank speeches on CBDCs collected at Auer et al (2020)¹. It contains 331 central bank speeches that explicitly mention CBDC from 2016 to 2022, from 44 different geographic areas. The collection of texts also includes the authors' expert judgment evaluation of the text's sentiment about CBDC. This will allow us to compare the expert's opinion (BIS² labeled data from now on) with that of the NLP methods.

NLP models: Dictionaries, BERT and ChatGPT

We employ three NLP techniques to measure sentiment on these speeches: Dictionaries, BERT and ChatGPT. Dictionaries can be considered as a benchmark. We use a popular dictionary in finance, Loughran and McDonald (2011), which is a predefined collection of words with positive and negative tone, to calculate the sentiment of a text by computing its polarity. BERT and ChatGPT are two of the most widely used LLMs. For BERT we implement 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 texts.

¹ Database updated up to January 2023.

² Bank for International Settlements.

ChatGPT, a version of the general GPT (Generative Pre-trained Transformer) models, is an auto-regressive Transformer model released by OpenAI in 2022. It has been trained on a much larger corpus than any other model up to 2023 (ChatGPT has been trained with 45 terabytes, compared to the 3 terabytes of BERT). To interact with ChatGPT it is necessary to specify the task to be carried out by means of a prompt. The selection of the prompt (known as prompt engineering) is very important, as different prompts could yield different results. Our benchmark prompt is:

*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: [...]*³

Workflow

Our text corpus is very heterogeneous, with different text sizes. The average size is around 3,500 words, and almost half of the documents have more than 3,000 words. While dictionary-based methods can be applied to texts of any size, both BERT and ChatGPT have limitations, of 500 tokens and 4,000 tokens respectively.⁴ Keeping in mind these limitations, we have decided to perform our analysis splitting the documents into paragraphs. Since the speeches cover many different topics other than CBDC, we select only relevant paragraphs where a number of keywords related to CBDC are mentioned. Once a relevant paragraph is selected, we calculate the sentiment with the three NLP methods. The overall sentiment of the document is then computed by averaging the sentiments of its paragraphs. We repeat this process for all the documents.

3. Comparing with human sentiment

Table 1 shows the correlation between the sentiment obtained by the three techniques with the sentiment labeled by human experts on Auer et al (2020), the BIS labeled data. The sentiment provided by each of the NLP methods is a continuous measure from -1 to 1, while the score from BIS labeled data is a discrete measure of -1, 0 or +1, depending on the document's sentiment towards CBDC. Results in **Table 1** are for all speeches, and for speeches with more than one relevant paragraph (larger texts, 242 out of 331). All correlations are positive and significant, which indicate that the techniques are capturing to some extent the sentiment expressed by humans. But LLMs are measuring better that sentiment, specially ChatGPT. Moreover, the difference between ChatGPT and the other techniques is bigger in larger texts, where the correlation between ChatGPT and BIS labeled data is significantly higher than the others.

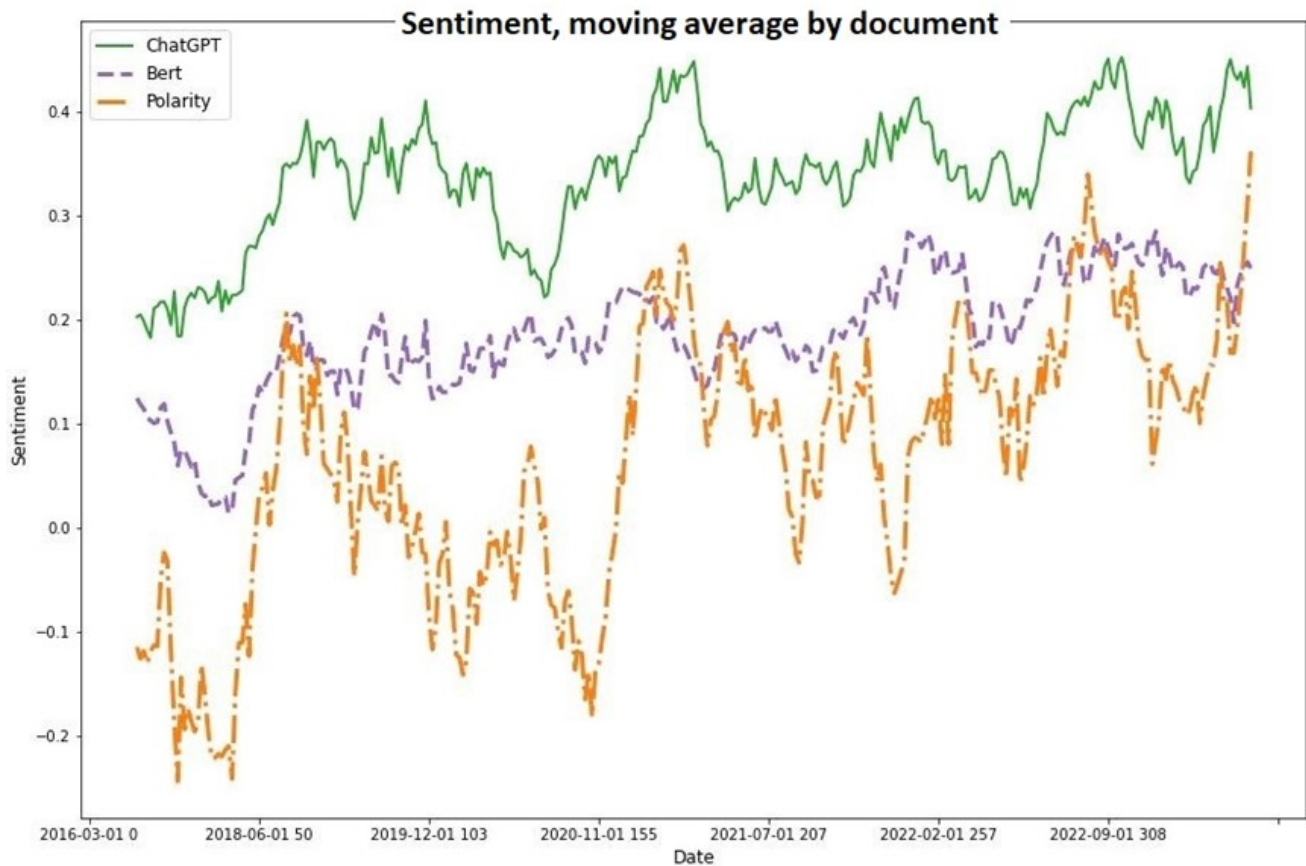
Table 1: Correlation with expert judgment

	All text	Larger texts
ChatGPT	0.34***	0.44***
BERT	0.29***	0.34***
Dictionary-based	0.21***	0.22***

Note: Correlation between the sentiment obtained by the three techniques with expert judgement, for all texts, and for larger texts.

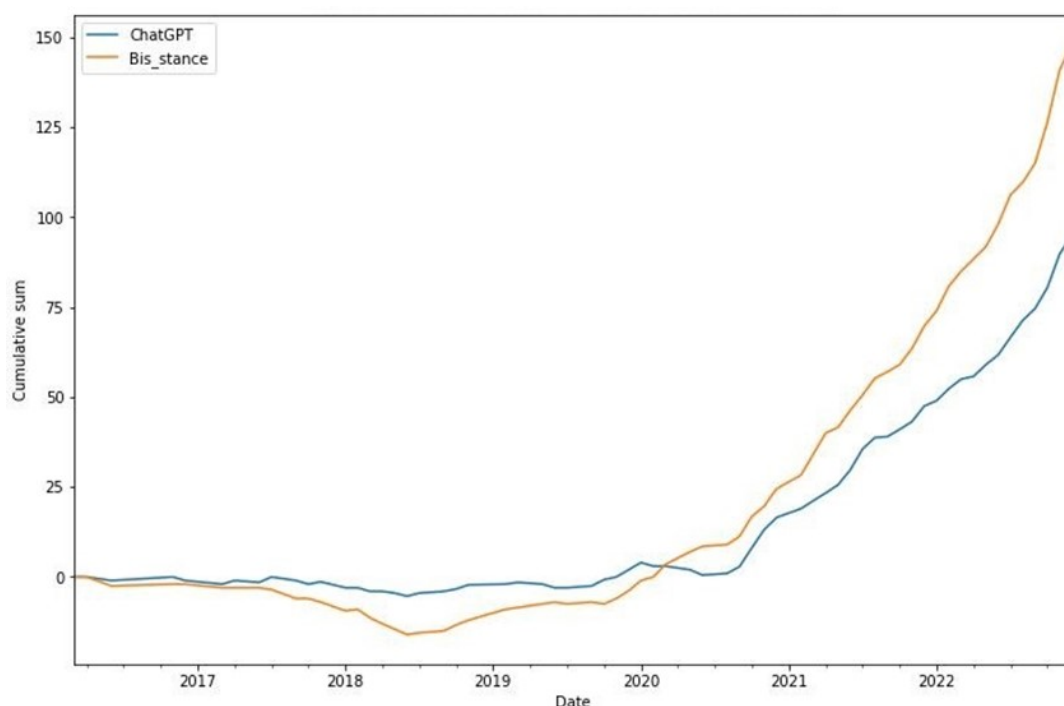
³ We have performed robustness analysis with different prompts. At the time of writing, the best version available in ChatGPT API was GPT-3.5, with a limitation of 4,000 tokens per prompt.

⁴ A token is a single unit of text, like words, numbers, or punctuation marks, separated by white space or other delimiters.



We now compare the temporal sentiment evolution of ChatGPT with labeled data BIS. 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 made by BIS). The result is shown in **Figure 1**, with BIS sentiment in orange and ChatGPT in blue. The trend changes are very similar, which is remarkable, especially considering that the prompt we used was quite generic, and it could be further tailored to capture BIS preferences.

Figure 1: Temporal evolution of sentiment by ChatGPT and expert judgement



Note: Comparing cumulative sum of ChatGPT (blue) and expert judgement (orange).

4. Conclusions and further work

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. 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.

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. 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).

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 our original article (Alonso-Robisco and Carbó 2023) we have carried out several experiments in this direction, with prompts that exclude CBDC from the input statement. 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. ■

References

- Alonso-Robisco, A. & Carbó, J. (2023). Analysis of CBDC Narrative of Central Banks using Large Language Models.
- Auer, Raphael, Giulio Cornelli and Jon Frost. (2020). "Rise of the central bank digital currencies: drivers, approaches and technologies".
- Auer, R., Frost, J., Gambacorta, L., Monnet, C., Rice, T., & Shin, H. S. (2022). Central bank digital currencies: motives, economic implications, and the research frontier. *Annual review of economics*, 14, 697-721.
- 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.
- McKay, A., Nakamura, E., & Steinsson, J. (2016). The power of forward guidance revisited. *American Economic Review*, 106(10), 3133-3158.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. arXiv preprint arXiv:1906.02243.
- Yang, Y., Uy, M. C. S., & Huang, A. (2020). Finbert: A pretrained language model for financial communications. arXiv preprint arXiv:2006.08097.

About the authors

Andrés Alonso Robisco joined Banco de España in 2019 as senior economist in the Financial Innovation Division where he analyses the latest trends in financial innovation. Specifically, he studies the impact of machine learning on credit risk modelling, and different topics related to climate finance innovation. Previously he had been working in the Single Resolution Board (SRB), an agency of the European Commission, and beforehand, in the treasury and capital markets teams of Instituto de Credito Oficial (ICO), the Spanish financial agency. He has published articles in several journals like *International Review of Financial Analysis*, and *Financial Innovation*.

José Manuel Carbó joined Banco de España in 2019 as a senior economist in the Financial Innovation Division, where he analyses the latest trends in financial innovation. Specifically, he studies the impact of machine learning on credit risk modelling, and different topics related to cryptocurrencies and financial stability. Prior to this, he was a consultant in ScanmarQED London and Research Associate in Imperial College London. He has a PhD in economics from Universidad Carlos III de Madrid. His research interests are machine learning, financial innovation, and policy analysis. He has published articles in several journals like *Journal of Applied Statistics*, *Economics of Transportation*, *Financial Innovation*, and *Regional Science and Urban Economics*, among others.

SUERF Publications

Find more **SUERF Policy Briefs** and **Policy Notes** at www.suerf.org/policynotes



SUERF is a network association of central bankers and regulators, academics, and practitioners in the financial sector. The focus of the association is on the analysis, discussion and understanding of financial markets and institutions, the monetary economy, the conduct of regulation, supervision and monetary policy.

SUERF's events and publications provide a unique European network for the analysis and discussion of these and related issues.

SUERF Policy Briefs (SPBs) serve to promote SUERF Members' economic views and research findings as well as economic policy-oriented analyses. They address topical issues and propose solutions to current economic and financial challenges. SPBs serve to increase the international visibility of SUERF Members' analyses and research.

The views expressed are those of the author(s) and not necessarily those of the institution(s) the author(s) is/are affiliated with.

All rights reserved.

Editorial Board

Ernest Gnan
Frank Lierman
David T. Llewellyn
Donato Masciandaro
Natacha Valla

SUERF Secretariat
c/o OeNB
Otto-Wagner-Platz 3
A-1090 Vienna, Austria
Phone: +43-1-40420-7206
www.suerf.org • suerf@oenb.at