ECBERT: Fine-Tuning ModernBERT for Monetary Policy Sentiment Analysis

Giovanni Raimondo Quaratino * Márton Barta † Agata Magdalena Dziwok ‡

Jan 2025

Keywords: Monetary Policy Sentiment Analysis, ModernBERT, GenAI, NLP

JEL classification: E58, C55, C63, G17

*European Central Bank. E-mail address: giovanni_raimondo.quaratino@ecb.europa.eu
†European Central Bank. E-mail address: marton.barta@ecb.europa.eu
‡European Central Bank. E-mail address: agata_magdalena.dziwok@ecb.europa.eu

1 Motivation

1.1 The Rising Importance of Central Bank Communication in Shaping Market Expectations

Central bank communications are increasingly recognized as pivotal for market participants, policymakers, and researchers who aim to gauge the direction of monetary policy. Whether in the form of policy statements, press releases, or speeches, these pronouncements can rapidly shift market expectations, influence asset prices, and even alter macroeconomic outcomes. Understanding the sentiment behind such communications—particularly the nuances that may be subtly encoded in tone or wording—has thus become a key challenge for the economic and financial community.

1.2 Leveraging Natural Language Processing to Decode Central Bank Sentiment at Scale

Recent advancements in Natural Language Processing (NLP) and machine learning present new opportunities for uncovering these sentiments at scale. Automated sentiment analysis tools can offer timely insights into how central bank messages are perceived by markets, potentially flagging shifts in policy stances that might otherwise go unnoticed until officially announced. Moreover, these techniques hold promise for enriching traditional macroeconomic models by incorporating textual data, thereby enhancing forecasts of inflation, interest rates, and overall economic activity.

1.3 Challenges and Opportunities in Advancing Sentiment Analysis of Monetary Policy

Yet, despite growing interest in this area, sentiment analysis of central bank communications remains an emerging field with significant room for methodological refinement. The specialized vocabulary, domain-specific context, and high-stakes implications of monetary policy require models that are both flexible and finely attuned to subtle textual cues. This evolving intersection of economics and AI underscores the importance of continued innovation—driven by ever-improving language models and increasingly granular data sources—to push the boundaries of how we interpret and act upon central bank communication signals.

2 Related Works

2.1 The Evolution of NLP in Economics: From BERT to Domain-Specific Models

The application of NLP to economics and finance has grown significantly in recent years, reflecting a broader trend of employing AI-driven techniques to extract signals from unstructured text. Among the pioneering approaches, Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. (2019)) has been lauded for its contextual bidirectional understanding of language, enabling superior performance in various text classification tasks. Building on BERT's success, RoBERTa (Liu et al. (2019)) introduced optimized training procedures and larger mini-batches to enhance model accuracy further. These innovations collectively fostered research on domain-specific language models capable of capturing the nuances in specialized corpora.

2.2 The Role of Central Banking Language Models

In the field of central banking and monetary policy analysis, researchers have begun to leverage these transformer-based architectures to interpret subtle shifts in tone and sentiment. Notably, the Central Banking Language Models (CB-LMs) introduced by the Bank for International Settlements (BIS) represent a cutting-edge effort to adapt BERT and RoBERTa for monetary policy sentiment analysis (Gambacorta et al. (2024)). By retraining on extensive datasets of policy communications and academic texts, then fine-tuning with annotated samples, the BIS successfully demonstrated the power of domain-focused models for detecting hawkish, dovish, or neutral stances in monetary statements.

2.3 ModernBERT: Advancing Sentiment Analysis with Enhanced Linguistic Sensitivity

A more recent breakthrough in model architecture is ModernBERT, whose enhancements to BERT's underlying layers and attention mechanisms reportedly achieve superior performance in tasks requiring high linguistic sensitivity (Warner et al. (2024)). This improvement has clear and important repercussions in central banking, where the differentiation between subtle rhetorical cues can have significant implications for financial markets. ModernBERT's ability to capture these fine-grained distinctions offers a new frontier for researchers seeking to improve the reliability and interpretability of sentiment classification in central bank communications.

3 Data and Methods

To refine ModernBERT into a model specifically created within the European Central Bank (ECB) for the task of monetary policy sentiment analysis (hence the name ECBERT), two main steps were undertaken: (1) assembling a specialized corpus for domain-specific pretraining, and (2) fine-tuning the resulting model for monetary policy sentiment analysis. Additionally, a comparative approach was used by creating multiple versions of the model to evaluate the impact of pretraining on downstream performance.

3.1 Data Collection and Pretraining

A custom corpus of **25,581 texts** was compiled for the pretraining stage (domain-adaptive Masked Language Modeling (MLM)). This corpus included:

- **498** texts from a dataset of European Central Bank **monetary policy decisions**.
- **22,604** speech **titles** from the BIS speeches dataset.
  - The decision to focus on titles rather than full speeches was made due to inconsistencies and unstructured formatting in the underlying text.
- **2,749** speech **titles** from the ECB speeches dataset.
  - As with the BIS speeches, only the titles were used because the full texts contained formatting issues and varied structure.

Following common practices in recent transformer-based research, **only MLM was used** for pretraining: the Next Sentence Prediction (NSP) task was omitted based on findings in Warner et al. (2024), which suggest that NSP often does not contribute significantly to downstream performance. Once assembled and preprocessed, **this entire pretraining corpus was uploaded to HuggingFace**[1], making it openly available for future research in monetary policy text analysis.

3.2 Fine-Tuning for Monetary Policy Sentiment

The primary goal of this study was to classify **monetary policy texts** as **"dovish", "hawkish", or "neutral".** To achieve this, we utilize the dataset from Gorodnichenko et al. (2023), where each sentence has been manually labeled by multiple domain experts to ensure high consistency. This dataset consists of historical Federal Open Market Committee (FOMC) statements from 1997 to 2010, which are divided into 1,243 sentences. Two fundamental moments are identified in the fine-tuning process:

1. **Data Splitting**
   - An **80-20** train-test split was applied to establish a baseline for model evaluation.

[1] `https://huggingface.co/datasets/Graimond/ECBERT-mlm-dataset`

- For more robust estimates, **30-fold cross-validation** was conducted, providing a comprehensive view of model performance across multiple subsamples of the data.

This approach follows the one employed by Gambacorta et al. (2024). In this regard, we also maintain the target label distribution within a 5 percentage point deviation. By identifying the top 30 datasets that best preserve this label distribution, it is possible to robsutly and reliably validating the findings.

2. **Model Variants**
   - **Fine-Tuned Only Model**: A baseline model in which the original ModernBERT architecture (without additional domain pretraining) was directly fine-tuned on the FOMC dataset.
   - **Pretrained-Only Model**: A version of ModernBERT pretrained on the 25,581 texts using MLM but not yet fine-tuned on the sentiment classification task.
   - **Pretrained + Fine-Tuned Model**: A fully adapted model that underwent the domain-specific MLM pretraining before being fine-tuned on the FOMC dataset.

By creating multiple variants, it was possible to isolate the effects of domain-specific pretraining from standard fine-tuning, shedding light on how each step contributed to classification accuracy.

3.3 Model Availability

Upon completion, **all three versions of the model**—the fine-tuned-only baseline[2], the pretrained-only model[3], and the pretrained + fine-tuned version[4]—were **released on HuggingFace**, ensuring transparency and reproducibility. This open-source approach aims to foster further experimentation and collaboration in the rapidly evolving area of monetary policy sentiment analysis.

[2] `https://huggingface.co/Graimond/ECBERT-base`

[3] `https://huggingface.co/Graimond/ECBERT-base-mlm`

[4] `https://huggingface.co/Graimond/ECBERT-base-pretrained-finetuned`

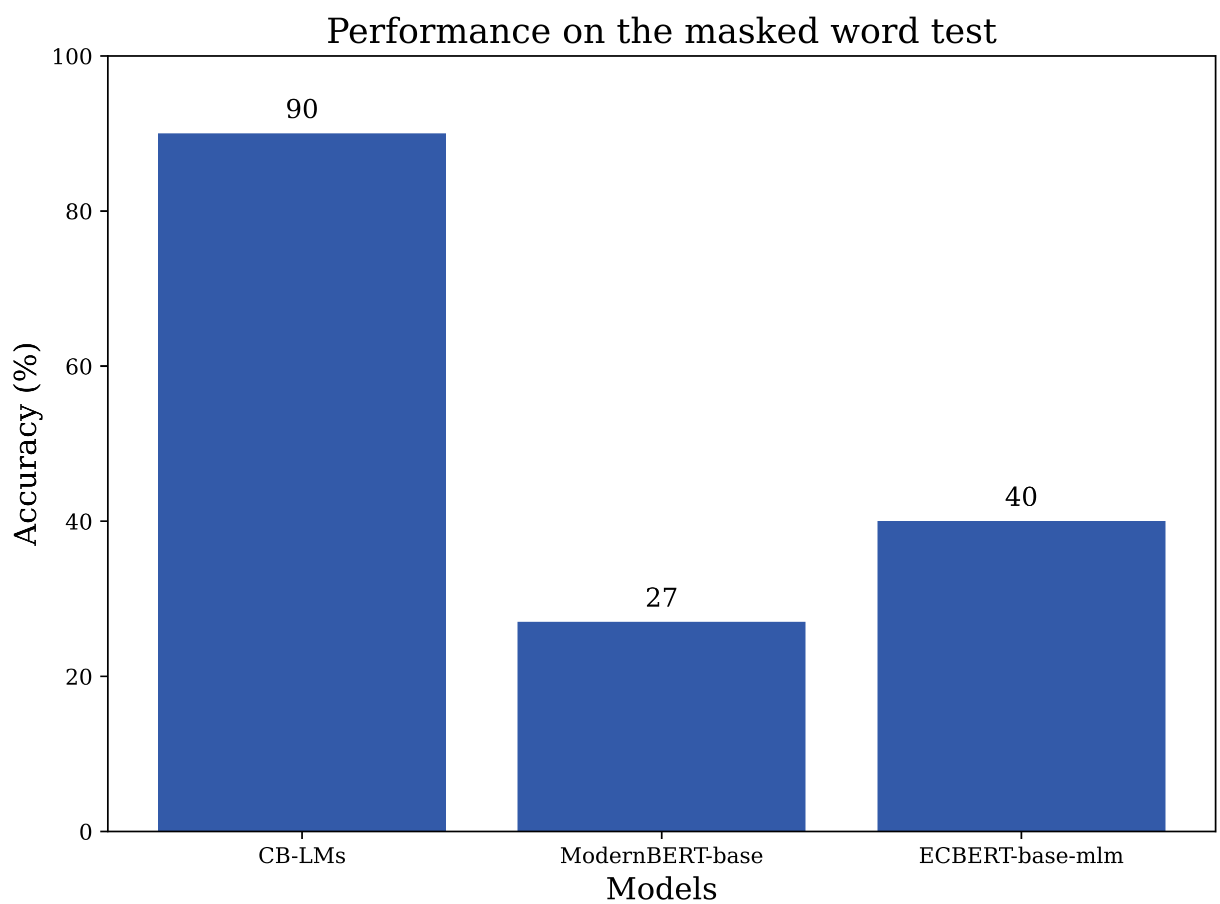
4 Results

This section presents the outcomes of two core stages: (1) Masked Language Modeling performance for domain-specific pretraining, and (2) monetary policy sentiment analysis accuracy following fine-tuning.

4.1 Masked Language Modeling Performance

The goal of the MLM task was to evaluate how effectively each model learned domain-specific vocabulary and context after being exposed to central bank communications. Three models were considered:

- **CB-LMs (BIS model)**: The best performing model by the BIS achieved a **90%** accuracy in predicting masked tokens.
- **ModernBERT-base**: When taken directly "out of the box" and tested on the same MLM task (i.e., without any domain adaptation), it reached only **27%** accuracy.
- **ECBERT-base-mlm**: After retraining on the custom corpus of 25,581 texts, its accuracy improved to **40%**.

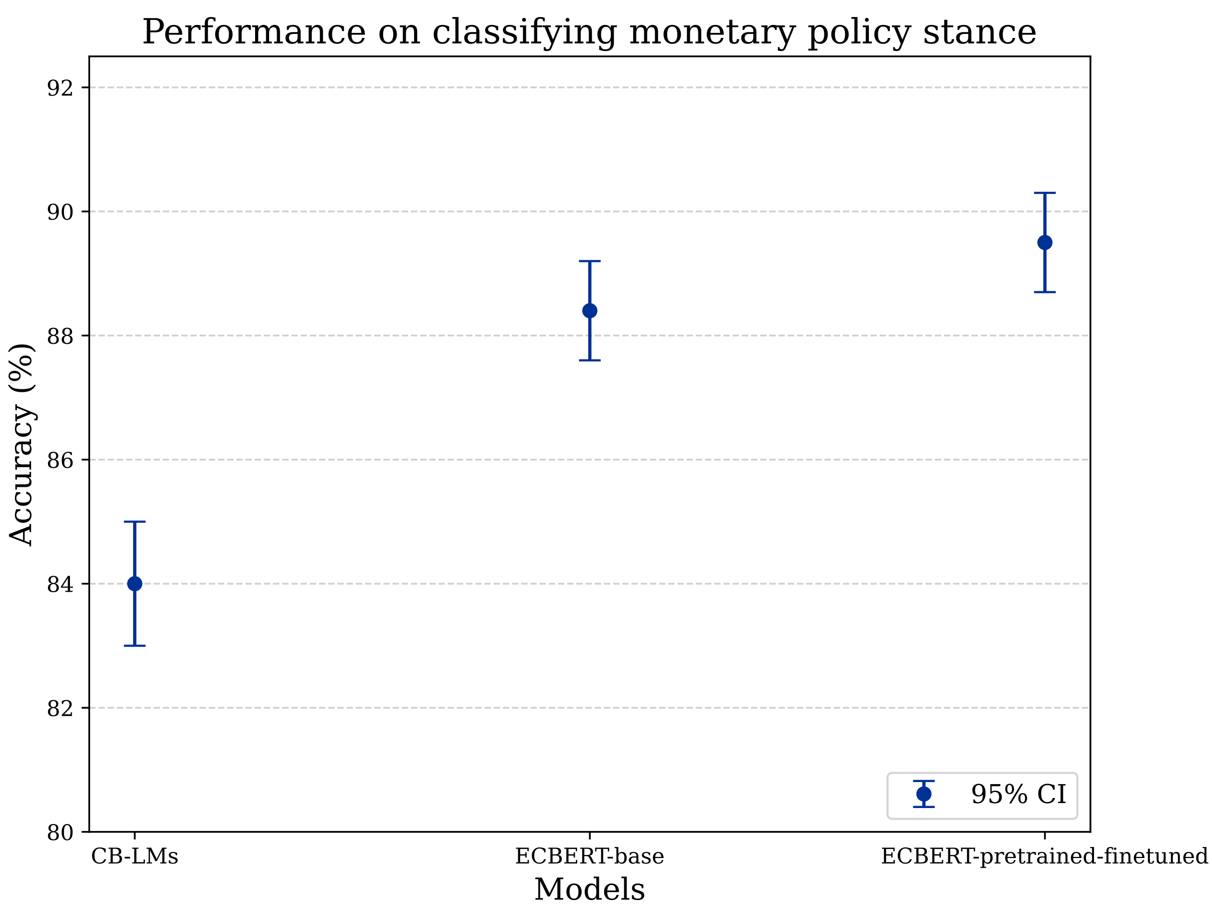


4.2 Monetary Policy Sentiment Analysis

For the downstream classification task—labeling texts as **dovish**, **hawkish**, or **neutral** three models are evaluated on the FOMC dataset. The final accuracy, averaged across **30** cross-validation splits, was as follows:

- **CB-LMs**: 84% average accuracy.
- **ECBERT-base**: 88.4% average accuracy.
- **ECBERT-base-pretrained-finetuned**: 89.5% average accuracy.

Here, too, additional domain-specific learning yielded meaningful gains: the ECBERT models surpassed the BIS CB-LMs, with the **pretrained + fine-tuned** variant performing best. These findings suggest that retraining ModernBERT on a curated central banking corpus before fine-tuning can substantially enhance a model's ability to capture subtle monetary policy sentiments. Overall, these improvements highlight the value of customization when deploying language models in specialized fields such as central banking.



References

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Burstein, J., Doran, C., and Solorio, T., editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Gambacorta, L., Kwon, B., Park, T., Patelli, P., and Zhu, S. (2024). CB-LMs: language models for central banking. BIS Working Papers 1215, Bank for International Settlements.

Gorodnichenko, Y., Pham, T., and Talavera, O. (2023). Data and code for: The voice of monetary policy.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. cite arxiv:1907.11692.

Warner, B., Chaffin, A., Clavié, B., Weller, O., Hallström, O., Taghadouini, S., Gallagher, A., Biswas, R., Ladhak, F., Aarsen, T., Cooper, N., Adams, G., Howard, J., and Poli, I. (2024). Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference.