

# Interdisciplinary Project - TU Vienna & WU Vienna

## Project Proposal - Understanding Central Bank Communication: A Machine Learning Analysis of ECB's Public Outreach

Paul von Hirschhausen (01453360)

az. Prof.in Dr.in Claudia Wutscher (Supervisor - WU Vienna)

Univ.Ass. Dr. Gábor Recski (Co-Supervisor - TU Vienna)

27.02.2025

## Contents

Abstract . . . . .	1
Motivation and Research Question . . . . .	2
Methodological Approach . . . . .	3
Glossary . . . . .	4
References . . . . .	4

## Abstract

This interdisciplinary project analyzes the European Central Bank's communication strategy through the lens of both legal technology and data science. Building on preparatory work completed during the "Fachseminar Legal Tech" at the WU Vienna, the project represents a collaboration with the WU Legal Tech Center. It employs state-of-the-art natural language processing techniques, specifically the BERTopic model, to analyze temporal patterns and thematic evolution in ECB's communication across different channels, with a particular focus on climate change and environmental concerns. The project combines advanced topic modeling with time series analysis to provide insights that bridge the gap between quantitative analysis of central bank communication and legal research. By analyzing both traditional press releases and blog articles, this research aims to contribute to the understanding of how the ECB's communication evolves across different platforms and time periods. The findings will serve as a valuable resource for ongoing legal research, particularly in relation to EU regulation commentary. The project leverages the expertise of the WU Legal Tech Center for validation of the topic modeling results, ensuring the reliable interpretation of legal and regulatory content within the analyzed communications. All code and documentation for this project can be found in the

GitHub repository at [https://github.com/Habenzu/ECB\\_Topic\\_Chronicles.git](https://github.com/Habenzu/ECB_Topic_Chronicles.git).

## Motivation and Research Question

The European Central Bank’s role encompasses both the supervision of credit institutions in the euro area and the implementation of monetary policy. Central to these responsibilities is their comprehensive communication strategy with various external stakeholders, including financial markets, political and institutional bodies, and the broader public. This fundamental aspect of their operations is explicitly outlined in their guiding principles:

“The communication policy of the ECB is an essential part of its accountability and good governance obligations as an independent monetary and supervisory authority. Regular contacts and interaction with members of the public, private sector, academia, interest groups, representative associations and civil society provide relevant input and information that help to understand the dynamics of the economy, the financial markets and the banking sector, and the broader societal context.” European Central Bank (2025)

Given the significance of ECB’s communication, researchers have extensively analyzed its various aspects and implications. Fortes & Le Guenedal (2021) examined its impact on financial markets, while Kaminskas & Jurkšas (2024) investigated the relationship between communication sentiment and monetary policy stance. Although both studies employed NLP techniques, particularly LDA and its variants for topic modeling, they limited their analyses to formal communication channels such as press conferences, monetary policy accounts, and official speeches by ECB Board members.

The proposed project aims to expand upon this existing research by incorporating ECB blog articles alongside traditional press releases, analyzing both through state-of-the-art topic modeling approaches. This enhanced dataset and modeling framework will address two primary research questions:

1. What temporal patterns emerge in ECB’s communication topics, and what meaningful differences exist between press releases and blog articles, as well as among different ECB Board members?
2. How has the prominence of climate change and environmental concerns evolved in ECB’s communication over time? Can we identify structural breaks in this topic’s distribution that indicate shifts in its relevance?

The findings from this project will serve as a supplementary resource for the revised edition of the Articles on Monetary Union in the so-called “Schwarze” commentary on the EU Treaties.

## Methodological Approach

This project comprises two main analytical components. The first component focuses on applying topic modeling to a dataset of ECB press releases and blog articles using BERTopic, an advanced topic modeling technique. BERTopic leverages transformer-based language models like BERT to generate document embeddings, subsequently applies dimensionality reduction to create manageable representations, clusters these reduced embeddings into dense groups, and finally employs c-TFIDF to identify significant terms (Grootendorst, 2022). Unlike traditional LDA approaches that rely on BoW or TF-IDF text representations, BERTopic’s use of pre-trained BERT embeddings enables it to capture both semantic relationships between words and their contextual usage (Kuo, 2023). This approach is expected to yield more accurate topic assignments per document, thereby creating a more reliable dataset for the second project component.

To establish a comprehensive understanding of topic modeling performance in the context of central bank communications, this project will include a systematic comparison between BERTopic and traditional Latent Dirichlet Allocation (LDA). While previous studies such as Kaminskas & Jurkšas (2024) and Fortes & Le Guenedal (2021) have primarily relied on LDA-based approaches, the relative performance of these methods specifically for ECB communications has not been thoroughly evaluated. Our comparison will assess both qualitative aspects (topic coherence, interpretability, and alignment with domain expertise) and quantitative metrics (coherence scores, silhouette scores for clustering quality, and computational efficiency).

To enhance the model’s performance, we will explore fine-tuning the embedding model within the BERTopic pipeline using our collected ECB communication data, aiming to better capture domain-specific semantic nuances. The quality of topic modeling will be validated through two approaches: manual verification by teaching assistants at the WU Legal Tech Center, and automated verification using a large language model.

The second component focuses on analyzing the temporal distribution of the identified topics to address our research questions. This analysis will employ various time series methodologies, including interrupted time series analysis and change point detection methods. Additionally, we will conduct descriptive analyses and create visual representations to explore and communicate the relationships between topics identified in the first component.

## Glossary

- BoW - Bag-of-Words
- c-TF-IDF - class-based TF-IDF
- ECB - European Central Bank
- LDA - Latent Dirichlet Allocation
- TF-IDF - Term Frequency and Inverse Document Frequency
- LDA - Latent Dirichlet Allocation

## References

- European Central Bank. (2025). *Guiding principles for external communication for high-level officials of the European Central Bank* [Policy Document]. European Central Bank. <https://www.ecb.europa.eu/ecb/our-values/transparency/html/eb-communications-guidelines.en.html>
- Fortes, R., & Le Guenedal, T. (2021). Tracking ECB’s communication: Perspectives and implications for financial markets. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3791244>
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv Preprint arXiv:2203.05794*.
- Kaminskas, R., & Jurkšas, L. (2024). ECB communication sentiments: How do they relate to the economic environment and financial markets? *Journal of Economics and Business*, 131, 106198. <https://doi.org/https://doi.org/10.1016/j.jeconbus.2024.106198>
- Kuo, C. (2023). *The handbook of NLP with gensim: Leverage topic modeling to uncover hidden patterns, themes, and valuable insights within textual data*. Packt Publishing. <https://books.google.at/books?id=LJPeEAAAQBAJ>