Discussant Remarks

From Text to Quantified Insights: A Large-Scale LLM Analysis of Central Bank Communication

Leif Anders Thorsrud

13th ECB Conference on Forecasting Techniques "Artificial intelligence in economic narratives, forecasting and risk assessments"

March 20, 2026

1. WHAT THE PAPER DOES

BI

Core idea

Transform central bank language into quantitative, comparable indicators of policy communication across countries and over time.

Data and method

- **Data:** 74,882 documents, 169 central banks, 1884–2025, about 21 million sentences.
- **Model:** fine-tuned multilingual sentence transformer.
- **Labels:** topic, communication stance, audience, and sentiment.

Main outputs

- Net policy sentiment
- Straightforwardness
- Explanation
- Net confidence

Each with forward- and backward-looking components

Why this is interesting

Novelty comes from the combination of **global coverage**, **multilingual text**, **sentence-level classification**, and **economically interpretable indices**.

2. MAIN FINDINGS

BI

- Inflation targeting changes communication. Central banks shift from backward-looking exchange-rate discussion toward forward-looking talk about inflation, interest rates, and activity.
- Forward-looking communication contains policy signals. The forward-looking component of net policy sentiment predicts future policy-rate moves and OIS rates; the backward-looking component mainly tracks current/past conditions.
- Audience targeting is systematic. Communication to the public is more confidence-building, to governments more risk-highlighting, and to markets more neutral/balanced.
- Communication style is state-dependent. Straightforwardness falls in crises; explanation rises in tightening phases.
- Risk communication matters too. Forward-looking confidence/risk language helps predict future market volatility.

Bottom line

The paper argues that central bank communication is not just narrative — it is an **active policy tool** operating through both monetary-policy and financial-stability channels.

3. OVERALL IMPRESSION

- Impressive data collection effort and scope
- Well-grounded overall contribution
  - Using a modern LM as an automated classification tool that systematically analyzes central bank communications along four key dimensions – topic, communication stance, audience, and sentiment – and thereby offers a comprehensive framework for evaluating policy messages.
- Earlier research has typically either studied only one dimension or used simpler methods – so important contextual nuances may be missing

4. COMMENTS

BI

1. Measurement and theory
2. About context and multilingual models
3. Classification structure and learning

4.1 MEASUREMENT AND THEORY

BI

Two of the indices speak directly to large and well-known research areas:

- The **net policy sentiment** captures the balance between hawkish and dovish signals, offering a quantitative measure of the directional tone of monetary policy communication.
- The **net confidence index** captures the balance between confidence-building and risk-highlighting language, offering distinctive insights when decomposed into forward- and backward-looking components.

The other two indices are more novel and sound reasonable, but I would like more discussion of the mechanisms linking them to economic outcomes:

- The **straightforwardness index** captures the extent to which monetary policy communication conveys unidirectional stance signals, distinguishing direct guidance from more conditional or hedged language
- The **explanation index** quantifies how central banks justify and contextualize policy decisions, capturing the narrative elaboration that accompanies different phases of the policy cycle.

In short, the breadth of the framework is a strength, but it also raises the burden of interpretation.

4.2 ABOUT CONTEXT AND MULTILINGUAL MODELS

BI

I find the argument that studying only one dimension (e.g., topics, sentiment, or communication stance) may miss important context convincing and plausible.

- Still, the focus is on sentence-level classification – might this still miss important context relative to, e.g., paragraph embeddings?

The large scope, both over time and across countries, is a major strength and clearly motivates a multilingual model.

- You test some robustness to this in Appendix B.
- To me, this choice goes directly to the paper's core contribution, and I would like to see more evidence on how the multilingual model performs relative to language-specific models – taking into account that the former can be fine-tuned on a large corpus, whereas the latter could only be trained on a smaller corpus.

4.3 CLASSIFICATION STRUCTURE AND LEARNING

The paper adopts a block factorization that assumes conditional independence between two pairs of labels:

$$P(y_{topic}, y_{stance}, y_{audience}, y_{sentiment}|x) \approx P(y_{topic}, y_{stance}|x)P(y_{audience}, y_{sentiment}|x)$$

because a full joint analysis would require an impractically large amount of labeled data, as the number of possible label combinations grows exponentially.

- I may be misunderstanding the fine-tuning setup, but does this effectively imply training two models?

Figure 7: Classes for the Topic and Communication Stance Dimensions

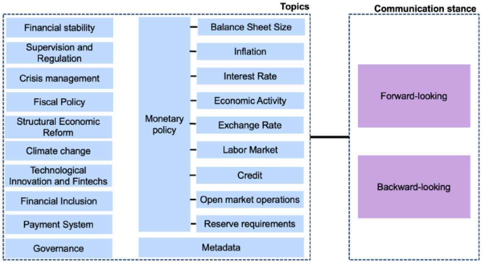
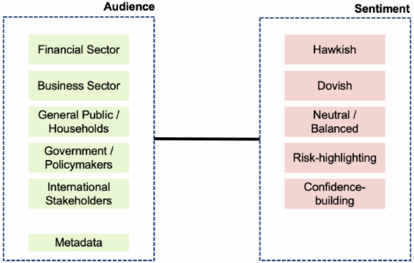


Figure 8: Classes for the Audience and Sentiment Dimensions



4.3 CLASSIFICATION STRUCTURE AND LEARNING CONT'D

Questions

- Would the framework perform worse if each model predicted only one label dimension (and not two)?
- From other settings we know that multitask learning often works well and improves overall loss - could that be an alternative to two models?

Again, to me these types of questions speak directly to the paper's core contribution: an automated classification tool that systematically analyzes central bank communications along four key dimensions – topic, communication stance, audience, and sentiment – and thereby offers a comprehensive framework for evaluating policy messages.

MINOR POINTS (FOR DINNER)

- A full section – Section 3 – is dedicated to "pro-forma analysis," focusing on lexical readability metrics and sentence complexity. While interesting, it may distract attention from the paper's main contribution.
- What about semantic drift or changes in meaning over time?

My takeaway

This work has a lot of potential: the data construction is impressive, the semantic measurement is genuinely useful, and the empirical results have clear policy relevance

What I find most convincing	74,882 documents, 169 central banks, 1884–2025, about 21 million sentences coupled with the fine-tuned multilingual LM
Most valuable next step	More exploration of model design and its consequences for learning well-defined monetary policy concepts
Why it matters	If the automated classification tool holds up, it provides a strong case for how multidimensional, context-sensitive learning can improve the measurement of policy communication