

Monetary Policy Communication According to Artificial Intelligence - Monetary Intelligent Language Agent (MILA)

Felix Geiger*
Dimitrios Kanelis*
Philipp Lieberknecht*
Diana Sola**

Large Language Models in Monetary Policy Research
Banque Centrale Du Luxembourg

22 January 2026

*Deutsche Bundesbank **DHBW Heidenheim

Disclaimer

The opinions expressed are those of the authors and do not necessarily reflect those of the Deutsche Bundesbank or the Eurosystem.

Motivation

- *Monetary policy is 98% talk and only 2% action* - Ben Bernanke (2015)
- Central bank communication affects expectations and stance (Clarida et al., 1999; Blinder et al., 2008; Blinder et al., 2024).
- Direct information about MP instruments & indirect information via CB's economic narrative.
- Research and markets analyze CB communication w.r.t. stance (hawkish/dovish) and sentiment (positive/negative).
- Due to this importance, policymaker need their own analysis tool, especially within the Eurosystem, given its cacophony of communication.

This Presentation: Policy and Technical Report from March 2025

- We develop a novel artificial intelligence (AI) methodology based on LLMs to evaluate MP texts in consistent, detailed, and transparent way
- **Monetary-Intelligent Language Agent (MILA):**
 - granular context-dependent analysis (Context Engineering)
 - Explainable AI
 - replicability
- We task MILA to evaluate the ECB's press conferences and speeches since Nov. 2011 (\approx 50,000 sentences)
 - dovish bias 2021-2022?
 - what happened in October 2022?

Related literature (1/2)

- Human Labeling (Jansen & de Haan, 2005; Ehrmann & Fratzscher, 2007; Rosa & Verga, 2007)
 - experts can consider context and detect nuances
 - subjectivity, inconsistency (time & context), time-consuming
- Word-frequency and Dictionary Algorithms (Apel & Blix Grimaldi, 2014; Picault & Renault, 2017, Aruoba & Drechsel, 2024)
 - consistency, transparency, speed
 - no context, mechanical, susceptible to nuances

Related literature (2/2)

- Predictive (Language) Models like BERT (Gorodnichenko et al., 2023; Curti & Kazinnik, 2023; Kanelis & Siklos, 2025)
 - consistency, can incorporate manual classifications
 - no context beyond text, fine-tuning costly and application-specific
 - static models that have problems when input data significantly deviates from the training data
- Plain LLM Usage (Hansen & Kazinnik, 2023; Gambacorta et al., 2025)
 - user-provided context and examples via prompt engineering, explanations
 - intransparent, stochastic, challenging to obtain detailed assessment

Methodology: Basics

- MILA combines text mining, a large language model, mathematical formulas and topic modeling.
- We build on the **paradigm shift** from fine-tuning to prompt and context engineering, which provides analytical flexibility for policy analysis.
- Role-based multi-layer prompt chaining with few-shot prompting in each layer (Wu et al., 2022; Sun et al., 2024).
- MILA builds on LLMs that we use with an OpenAI-compatible Python API:
 - This Presentation: **Llama 3.1 70B**
 - Currently in Practice: **Gemma 3 27B**
 - Planned: **GPT-5**

Methodology: Decision Hawk-O-Meter

- **Objective:** Construct an indicator ranging from -10 to +10 to quantify the stance of **direct interest rate communication**.
- **Key Challenge:** How can we be confident that the resulting score **is not merely a hallucinated number** produced by the model?
- **Proposed idea:** Use a **parallel prompt chain** consisting of four distinct sub-indicators, which are combined into a composite measure.
- **„Needle in the Haystack“- approach:** Identify the specific sentences relevant for answering the question and derive the numerical score directly from those answers, rather than generating the number in one step.

Methodology: Decision Hawk-O-Meter

- **Interest Rate Score:** Will key interest rates be lowered, raised, or maintained, and if changed, how many basis points? $\{-0.3, -0.15, 0, 0.15, 0.3\}$
- **Interest Rate Outlook Score:** Are signals for future reductions or increases communicated? Is the language cautious, neutral, data-dependent, or clearly directive, or is it subject to directional distortions? $\{-0.3, -0.15, 0, 0.15, 0.3\}$
- **Inflation Score:** Are the decision justified by direct reference to inflation developments, and is there an indication of upward or downward risk? $\{-0.2, -0.1, 0, 0.1, 0.2\}$
- **General Tone Score:** What is the tone of the language, and what other announcements are made with implications for the interest rate path? Is the language particularly cautious, supportive, expansionary, or markedly restrictive? $\{-0.2, -0.1, 0, 0.1, 0.2\}$

Example: Interest Rate Outlook Score

- **Interest Rate Outlook Score:** Are signals for future reductions or increases communicated? Is the language cautious, neutral, data-dependent, or clearly directive, or is it subject to directional distortions?
 - Clear commitment to future rate hikes or monetary tightening. → 0.30
 - Subtle hints toward possible tightening measures or explicit announcements to maintain currently restrictive policy levels for a longer time. → 0.15
 - Data-dependent, neutral stance (no clear directional guidance). → 0
 - Subtle hints toward possible easing measures. → -0.15
 - Clear commitment to future rate cuts or monetary easing or explicit announcements of a long-lasting expansionary policy stance. → -0.30

Methodology: Narrative Hawk-O-Meter and Sentiment

- We define euro-area hawkish/dovish and macro positive/negative tones based on literature and institutional knowledge, then generate corresponding prompts with few-shot examples.
- **Prompt Chaining:** One LLM layer derives the inflation context from ECB/Eurosystem staff projections and another the intended monetary stance as a score from ECB monetary policy decisions. → **Input for the next LLM layer.**
- **Explainability:** Classification on sentence level in hawkish/dovish & positive/negative, taking document context and inflation context into account.
 - **Prompt Chaining:** One LLM layer classifies sentences as hawkish or dovish, and the next layer determines whether they are moderate or not.
- Calculation of hawkometer and sentiment for overall text and topics using different mathematical formulas, with baseline formulas:

$$Sentiment_{i,t} = \frac{\#Positive_{i,t} - \#Negative_{i,t}}{\#Positive_{i,t} + \#Negative_{i,t}} \quad Hawkometer_{i,t} = \frac{\#Hawkish_{i,t} - \#Dovish_{i,t} + \frac{1}{2}(\#ModerateHawkish_{i,t} - \#ModerateDovish_{i,t})}{\#Hawkish_{i,t} + \#Dovish_{i,t} + (\#ModerateHawkish_{i,t} + \#ModerateDovish_{i,t})}$$

Prompting Details for the Narrative Hawk-O-Meter (1/2)

Hawkish refers to the communication of information that either justifies tighter monetary policy or makes future monetary policy tightening more likely. Such sentences point to upside inflation risks arising from:

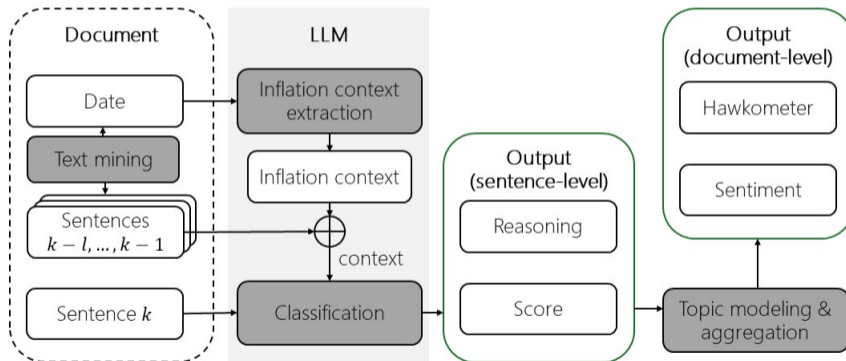
- rising inflation or mounting inflation pressures,
- economic developments that increase purchasing power,
- adverse supply shocks (e.g., higher oil or food prices, trade disruptions, or shortages),
- tight labor markets or strong wage growth,
- overheating or unsustainable high economic activity,
- rapid or unsustainable growth in money or credit,
- criticism of expansionary public spending or high public debt, and support for policies aimed at reducing inflation.

Prompting Details for the Narrative Hawk-O-Meter (2/2)

Dovish refers to the communication of information that either justifies easier monetary policy or makes future monetary policy easing more likely. Such sentences point to downside inflation risks arising from:

- declining inflation or deflation risks,
- economic developments that reduce purchasing power,
- increasing unemployment or muted wage growth,
- weak economic activity,
- slow or weak growth in money or credit,
- supporting expansionary public spending,
- risks to financial stability.

Methodology: Multi-Layer Prompt Chaining



Example: Context Relevance and Counter Factual from April 2025

- A new sentence from April 2025: „**A boost in defence and infrastructure spending could also raise inflation through its effect on aggregate demand.**“
- Given the actual context of inflation risks being tilted to the upside, MILA evaluates:
- **Stance:** *Hawkish* and **Sentiment:** *Negative*
- Assuming this sentence was communicated during a low-inflation period (below target), MILA evaluates:
- **Stance:** *Moderate Hawkish* and **Sentiment:** *Positive*

Some First Evaluation based on Individual Sentences

Evaluation				
Metric	Sentiment		Hawk-O-Meter	
	LM (2011)	MILA	AB (2014)	MILA
Accuracy	0.76	0.96	0.43	0.84
Precision	0.61	0.90	0.67	0.84
Recall	0.97	0.99	0.43	0.84
F1-Score	0.75	0.94	0.35	0.83
Data	Pfeifer and Marohl (2023) ECB speeches		Nitoi et al. (2023) EE CB minutes	
Sentence Context	NO		NO	

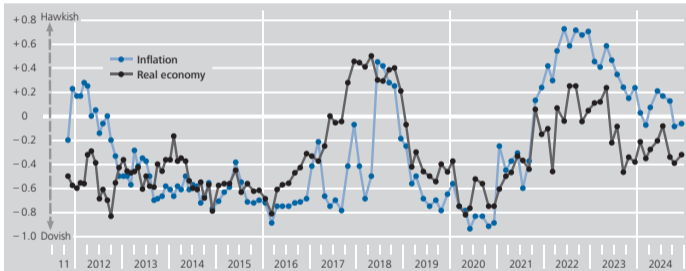
Work-in-Progress: A proper framework and data set to evaluate context-dependent classification for euro area CB communication.

Replicability: MILA's iteration correlation is ≈ 0.98 , despite the stochastic nature of LLMs (full document approach as low as 0.6)

Hawk-O-Meter for ECB Press Conferences

Hawk-O-Meter of economic narratives in ECB press conferences*

Chart 3.X



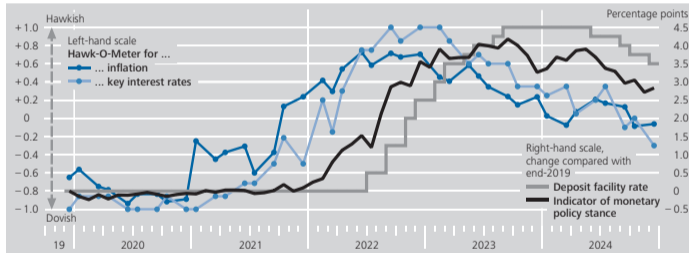
Source: Bundesbank analysis using MILA, an AI economist based on Llama 3.1. * The Hawk-O-Meter measures whether communication is indicative of restrictive (hawkish) or accommodative (dovish) monetary policy. The points represent classifications of individual monetary policy statements. Deutsche Bundesbank

- Inflation narrative in 2022–2023 was distinctly hawkish, peaking with the first rate hike in July 2022
- Communication on real economy more dovish: supply-side disruptions, weak growth

Was There a Dovish Bias in 2021–2022?

Hawk-O-Meter for ECB press conferences compared with changes in key interest rates*

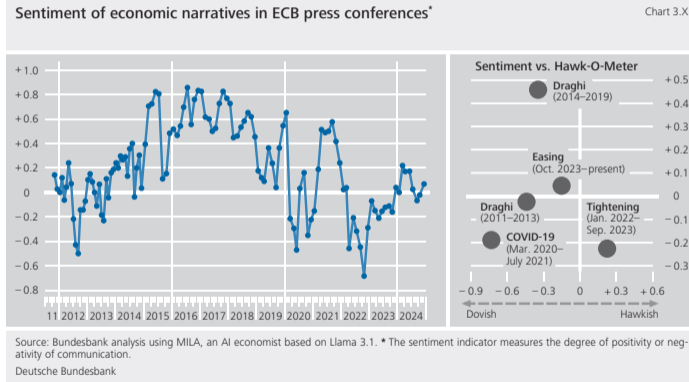
Chart 3.X



Source: Bundesbank analysis using MILA, an AI economist based on Llama 3.1. * The Hawk-O-Meter measures whether communication is indicative of restrictive (hawkish) or accommodative (dovish) monetary policy. The points represent classifications of individual monetary policy statements. Deutsche Bundesbank

- Consistently dovish communication on interest rates and inflation during the COVID-19 pandemic
- Late 2021: Qualitative divergence between communication on interest rates and inflation
- Communication influenced monetary policy stance even before the first interest rate hike

Sentiment of ECB Press Conferences

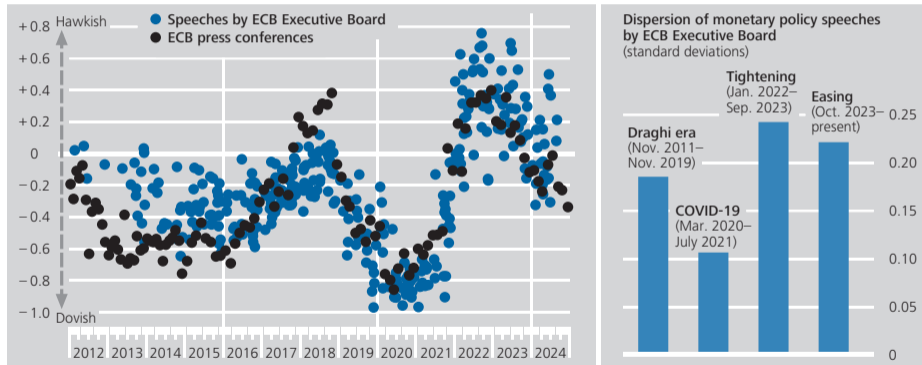


- Economic optimism and pessimism reflected in ECB communication.
- Pessimistic tone during the sovereign debt crisis, the pandemic, and the tightening phase.
- Particularly notable: **Press conference in October 2022**

Hawk-O-Meter of monetary policy speeches

Hawk-O-Meter for monetary policy speeches and ECB press conferences*

Chart 3.X



Source: Bundesbank analysis using MILA, an AI economist based on Llama 3.1. * The Hawk-O-Meter measures whether communication is indicative of restrictive (hawkish) or accommodative (dovish) monetary policy.

Deutsche Bundesbank

Hawk-O-Meter Taylor Rule for ECB Speeches (1/2)

$$\begin{aligned} \text{Hawkometer}_{t,i} = & \beta_0 + \beta_1 \cdot \text{HICP Inflation}_{t,i} + \beta_2 \cdot \text{GDP Growth}_{t,i} \\ & + \alpha_i + \gamma_t + \epsilon_{t,i} \end{aligned}$$

- *Hawkometer*: Hawkometer score for speaker i 's speech at time t
- *HICP Inflation*: Three-month euro area average of HICP inflation prior to the speech
- *GDP Growth*: Three-month euro area average of GDP growth prior to the speech
- α_i : Speaker Fixed-Effect
- γ_t : Time Fixed-Effect

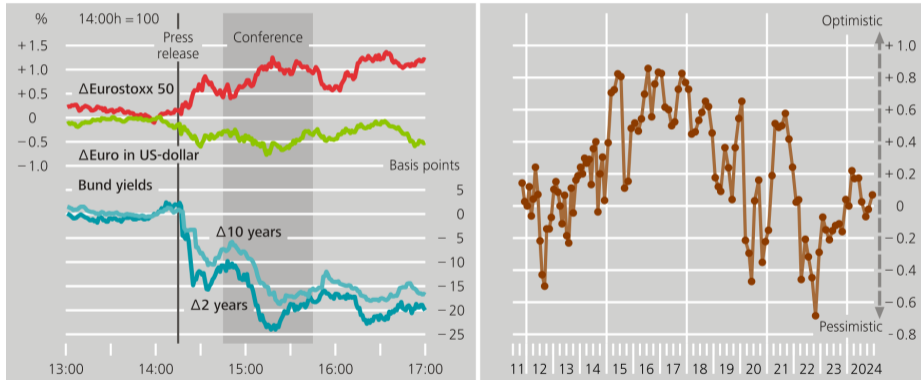
Hawk-O-Meter Taylor Rule for ECB Speeches (2/2)

Variable	Hawkometer ECB speeches				
	(1)	(2)	(3)	(4)	(5)
HICP inflation	0.095*** (0.005)	0.130*** (0.005)	0.064*** (0.010)	0.068*** (0.009)	0.067*** (0.009)
GDP growth	0.015*** (0.004)	0.002 (0.003)	0.012*** (0.005)	0.011*** (0.004)	0.009** (0.004)
Fixed effects	-	Speaker	Year	Speaker, Year	Speaker \times Year
N	350	350	350	350	350
R^2	0.57	0.72	0.78	0.83	0.85

Note: Standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Monetary policy speeches from Nov. 2011 - Aug. 2024. Regressors are three-month euro area averages.

What happened in October 2022?

Market Reaction of ECB Press Conference from 27th October 2022



Source: Bundesbank analysis using MILA, an AI model based on Llama 3.1.
Deutsche Bundesbank

6. Okt. 2025, 09:50:49, Vo4X0159_PRb(engl)

Conclusion and Policy Experiences

- MILA combines advantages of previous methods: state-dependent, granular, transparent, flexible, consistent
⇒ quasi-explainable, robust and detailed quantitative indicators for properties of central bank communication.
- Future avenue: creating an evaluation framework and data set for euro area central bank communication; compare different LLMs within the MILA framework.
- Regular use in ex-post review of ECB meeting (and inter-meeting) communication and ex-ante analysis of selected speeches by Bundesbank governor.
- Ongoing policy work: extension to post-meeting Q&A sessions, other speeches and meeting accounts.

Thank you!

Questions?

References

- Apel, M., Blix Grimaldi, M., 2014. *How Informative Are Central bank Minutes?* Review of Economics 65
- Aruoba, S. B. and T. Drechsel (2024), *Identifying Monetary Policy Shocks: A Natural Language Approach*, NBER Working Paper No 32417, May 2024.
- Blinder, A. S., M. Ehrmann, M. Fratzscher, J. De Haan and D.-J. Jansen (2008), Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence, Journal of Economic Literature, Vol. 46, No 4, pp. 910-945.
- Blinder, A. S., M. Ehrmann, J. De Haan and D.-J. Jansen (2024), Central Bank Communication with the General Public: Promise or False Hope, Journal of Economic Literature, Vol. 62, No 2, pp. 425-457.
- Clarida, R., J. Galí and M. Gertler (1999), The Science of Monetary Policy: a New Keynesian Perspective, Journal of Economic Literature, Vol. 37, No 4, pp. 1661-1707.
- Curti, F. and S. Kazinnik (2023), *Let's face it: Quantifying the impact of nonverbal communication in FOMC press conferences*, Journal of Monetary Economics, Vol. 139, pp. 110-126

References

- Deutsche Bundesbank, 2025. *Monetary Policy Communication According to AI*. Monthly Report March 2025
- Ehrmann, M., Fratzscher, M., 2007. *Communication by Central Bank Committee Members: Different Strategies, Same Effectiveness?* Journal of Money, Credit, and Banking 39, 509–541
- Gambacorta, L., Kwon, B., Park, T., Patelli, P., Zhu, S., 2024. *CB-LMs: Language Models for Central Banking*. BIS Working Papers
- Gorodnichenko, Y., T. Pham and O. Talavera (2023), *The Voice of Monetary Policy*, American Economic Review, Vol. 113, No 2
- Hansen, A. L. and S. Kazinnik (2024), *Can ChatGPT Decipher FedSpeak?*, Working Paper
- Jansen, D.-J. and J. De Haan (2005), *Talking heads: the effects of ECB statements on the euro-dollar exchange rate*, Journal of International Money and Finance, Vol. 24, No 2, pp. 343-361

References

- Kanelis, D., Siklos, P.L., 2025. *The ECB Press Conference Statement: Deriving a New Sentiment Indicator for the Euro Area*. International Journal of Finance & Economics, 652–664
- Loughran, T., McDonald, B., 2011. *When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks*. Journal of Finance 66(1), 35–65
- Nitoi, M., Pochea, M.M., Radu, S.C., 2023. *Unveiling the sentiment behind central bank narratives: A novel deep learning index*. Journal of Behavioral and Experimental Finance 38
- Pfeifer, M., Marohl, V.P., 2023. *CentralBankRoBERTa: A fine-tuned large language model for central bank communications*. The Journal of Finance and Data Science 9, 2405–9188
- Picault, M., Renault, T., 2017. *Words are not all Created Equal: A New Measure of ECB Communication*. Journal of International Money and Finance, 136–156
- Rosa, R., Verga, G., 2007. *On the consistency and effectiveness of central bank communication: Evidence from the ECB*. European Journal of Political Economy, 23(1), pp. 146-175

Literature

- Sun, S., Yuan, R., Cao, Z., Li, W., Liu, P., 2024. *Prompt Chaining or Stepwise Prompt? Refinement in Text Summarization*. Accepted to Findings of ACL 2024
- Wu, T., E. Jiang, A. Donsbach, J. Gray, A. Molina, M. Terry and C. J. Cai (2022), *PromptChainer: Chaining Large Language Model Prompts through Visual Programming*, arXiv 2203.06566.