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

Deutsche Bundesbank
Felix Geiger
Dimitrios Kanelis
Philipp Lieberknecht
Diana Sola

The Impact of AI on Economy, Finance and Supervision 2025

6th November 2025

Motivation I

- *"Monetary policy is 98% talk and only 2% action"* - Ben Bernanke (2015)
- Central bank (CB) communication affects expectations and monetary policy (MP) stance (Clarida et al., 1999; Blinder et al., 2008; Blinder et al., 2024).
- Communication contains direct information about MP instruments & indirect information via CB's economic narrative.
- Academic research and financial markets analyze CB communication with respect to stance (hawkish/dovish), sentiment (positive/negative) and other dimensions.

Motivation II

- Readily available AI tools capable of digesting large amounts of text quickly.
- Financial market participants increasingly use such tools to analyze policy texts.

1. How do these AI tools assess policy texts and on what grounds?
2. What are the implications of such AI-based assessments for market expectations, asset prices and financial stability?
3. How should policymakers react to widespread AI-based assessments?

$\Rightarrow$ To answer these questions, policymakers need their own AI tools.

Our contribution

- We develop a novel artificial intelligence (AI) methodology based on large language models (LLMs) to evaluate MP texts.

- **Monetary-Intelligent Language Agent (MILA)**:
  - granular
  - context-dependent (Context Engineering)
  - flexible
  - transparent (Explainable AI)
  - replicable

- We task MILA to evaluate ECB press conferences and speeches since Nov. 2011 ($\approx$ 50,000 sentences).
  - Case study 1: Was there a dovish bias 2021-2022?
  - Case study 2: What happened at the ECB press conference 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

  $\Rightarrow$ static models that struggle when input deviates significantly from 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

  $\Rightarrow$ black-box models that use uncertain context for their tasks

Methodology (1/3)

- We build on the paradigm shift in Natural Language Processing (NLP) from fine-tuning to prompt and context engineering, which provides analytical flexibility.
- MILA combines text mining, a large language model, mathematical formulas and topic modeling.
- MILA is based 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**
- We use role-based multi-layer prompt chaining with few-shot prompting in each layer (Wu et al., 2022; Sun et al., 2024) for classifying individual sentences.

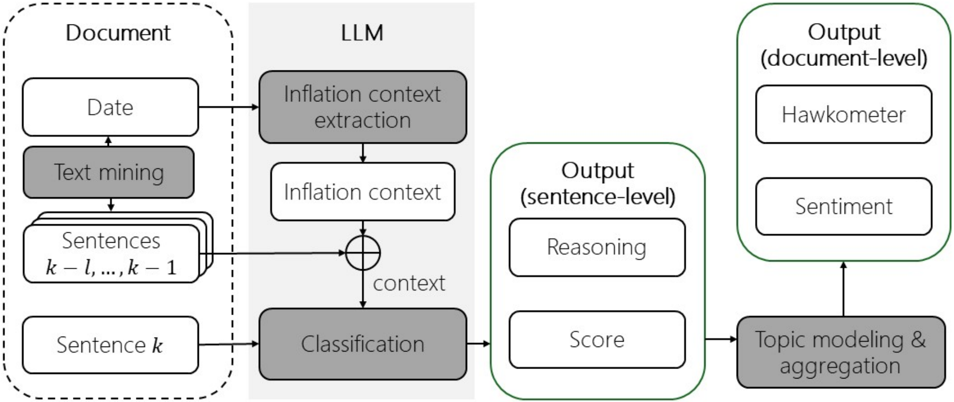
Methodology (2/3)

- **Role-based:** Assign LLM role of MP expert equipped with definitions of hawkish/dovish and positive/negative based on research literature and EA institutional knowledge.
- **Multi-layer prompt chaining:** Classification task divided across sequential LLM layers.
  1. Derive inflation context from ECB/Eurosystem staff projections
  2. Derive intended monetary policy stance from ECB decisions
  3. Classify individual sentences in hawkish/dovish and positive/negative based on 1 & 2.
- **Few-shot prompting:** Each layer receives example solutions for its respective task.
- **Aggregation:** Calculate overall score for text using pre-defined formulas from literature.

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

- **Topic modeling:** Assign topic to individual sentences to allow calculating topical scores.

Methodology (3/3)



Example: Context relevance and counterfactual from April 2025

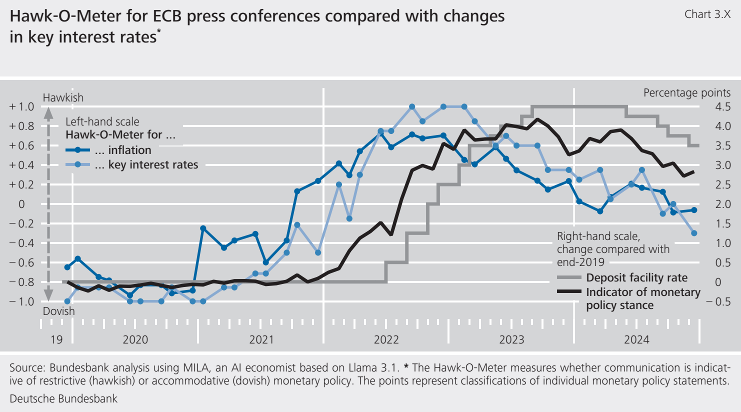
- **„A boost in defence and infrastructure spending could also raise inflation through its effect on aggregate demand.“**
- Given the context of inflation risks being tilted to the upside, MILA evaluates:

  Stance: *Hawkish* and **Sentiment:** *Negative*

- Assuming this sentence was communicated during a below-target period, MILA evaluates:

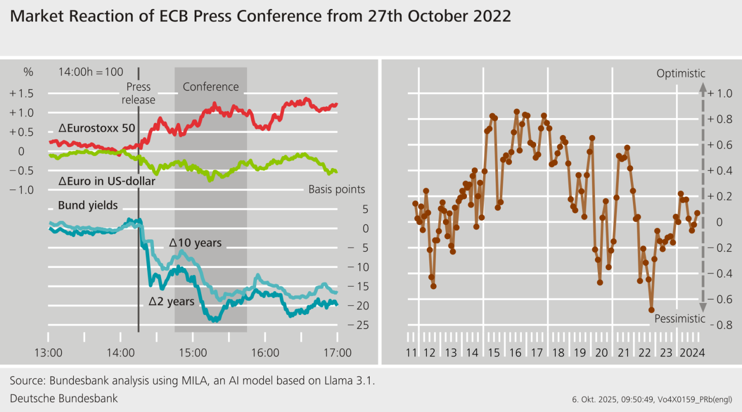
  Stance: *Moderately Hawkish* and **Sentiment:** *Positive*

Case study 1: Was there a dovish bias in 2021–2022?



- Consistently dovish communication during COVID-19 pandemic
- Late 2021: Tones on interest rates and inflation diverge
- Communication influenced stance even before first hike

Case study 2: What happened at the ECB press conference in Oct. 2022?



- 2Y Bunds declined by 20 bps despite policy rate hike by 75 bps
- MILA: sentiment in Policy Statement extremely negative
- Informative for shock identification and policy design

Conclusion

- MILA combines advantages of previous methods: quasi-explainable, robust and detailed quantitative indicators for properties of central bank communication.
- AI tools are able to assess policy texts quickly, but sensible results require careful design; quick-and-dirty approaches may be misleading.
- Carefully designed AI tools can be useful for policymakers ex ante and ex post.
  - At Bundesbank, MILA is used regularly to review ECB communication and to cross-check selected speeches by Governor before publication.
- Essential that policymakers are aware of the type of AI tools used by financial market participants to analyze policy texts and communication.

Thank you!

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.

Motivation

- Narrative approach: "gather systematic evidence from contemporaneous qualitative sources [...] to achieve macroeconomic identification" (Romer & Romer, 1989)

- "[T]he main way to use the narrative approach is going to be the old-fashioned one — scholars sitting at their desks [...] doing a lot of careful reading of the narrative sources themselves [...] we thoroughly expect to be made largely redundant by computers eventually, but perhaps not for a few years to come" (Romer & Romer, 2023)

How MILA analyzes a ECB monetary policy statement

1. Derive medium-term inflation context (overall text)
2. Analyze monetary policy decisions (1st part of MPS)

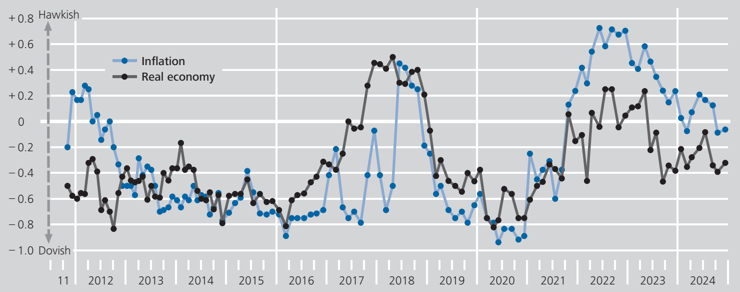
   $\Rightarrow$ decision hawkometer
3. Analyze economic narrative (individual sentences) based on step 1 and 2
   - hawkometer classification (2nd part of MPS) in two layers
   - sentiment classification (overall text)
4. Calculate overall hawkometer and sentiment score for economic narrative
5. Differentiate between inflation and real economy hawkometer/sentiment

Classification examples

- The disinflation process is well on track. (Positive)
- The Governing Council expresses its full support to the people of Ukraine. (Neutral)
- The Russian invasion of Ukraine is a watershed for Europe. (Negative)
- A boost in defence and infrastructure spending could also raise inflation through its effect on aggregate demand. (Hawkish)
- Inflation is expected to rise in the coming months, partly because previous sharp falls in energy prices will drop out of the annual rates. (Moderately hawkish)
- Most measures of longer-term inflation expectations stand at around 2 per cent. (Neutral)
- Manufacturing is still a drag on growth even if survey indicators are improving. (Moderately dovish)
- Annual inflation fell further to 1.7 per cent in September, its lowest level since April 2021. (Dovish)

Hawk-O-Meter for ECB Press Conferences

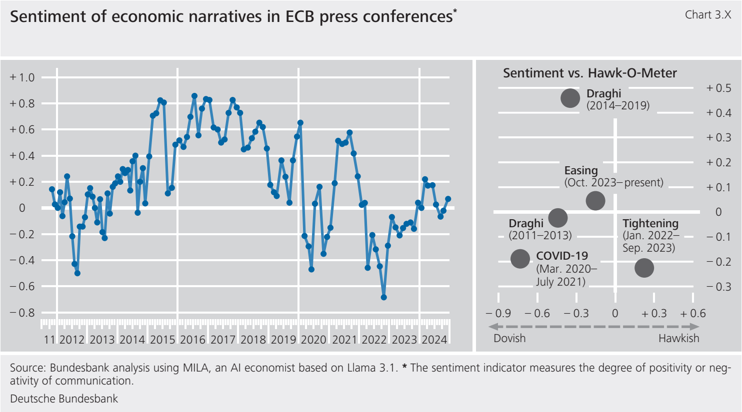
Hawk-O-Meter of economic narratives in ECB press conferences* Chart 3.X



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

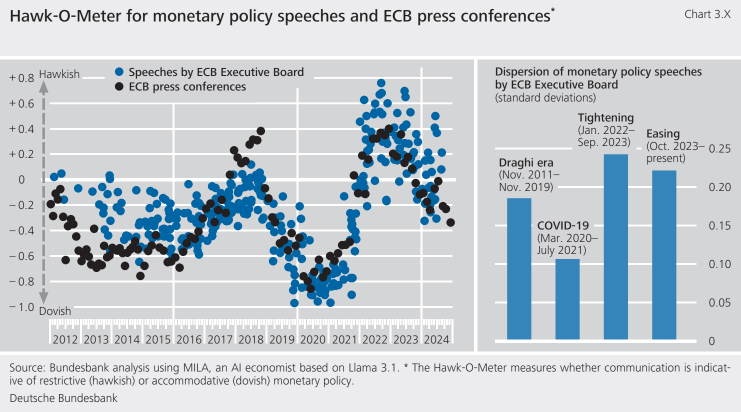
- Inflation narrative in 2022–2023 distinctly hawkish, peaking with first hike in July 2022
- Communication on real economy more dovish: supply-side disruptions, weak growth

Sentiment of ECB Press Conferences



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

Hawk-O-Meter of monetary policy speeches



- Low dispersion of Executive Board speeches during pandemic.
- Dispersion more pronounced than historically during 2022-2023 tightening phase and subsequent easing phase.

Hawk-O-Meter Taylor Rule for ECB Speeches

$$\begin{aligned} Hawkometer_{t,i} = \beta_0 + \beta_1 \cdot HICP\ Inflation_{t,i} + \beta_2 \cdot GDP\ Growth_{t,i} \\ + \alpha_i + \gamma_t + \epsilon_{t,i} \end{aligned}$$

- *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

Case study 3: Was ECB communication in line with a Taylor rule?

	Hawkometer ECB speeches				
Variable	(1)	(2)	(3)	(4)	(5)
HICP inflation	0.095***	0.130***	0.064***	0.068***	0.067***
	(0.005)	(0.005)	(0.010)	(0.009)	(0.009)
GDP growth	0.015***	0.002	0.012***	0.011***	0.009**
	(0.004)	(0.003)	(0.005)	(0.004)	(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.

Evaluation based on Individual Sentences

Evaluation

	Sentiment		**Hawk-O-Meter**	
Metric	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 dataset 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)

Technical infrastructure

- LLM on-premise in private cloud operated by Bundesbank (OPALS, On-Premise Allocated Language Model Service)
  - data security, flexibility, cost-efficiency
  - prototype by SCAI (IT-7)
- Python API (application programming interface) to LLM
  - crucial for efficient application of multi-layer prompt strategy
  - possibility to set model parameters, e.g. temperature

Risks of AI analyses in financial markets

- Growing use of AI to analyze monetary policy may homogenize interpretations across market participants.
- Reliance on similar AI models and AI assessments reduces incentives to gather diverse, additional information.
- Incentives to anticipate others' views encourage alignment with AI-based evaluations (herding behavior).
- Expectation formation and price discovery become less efficient; market sentiment may turn fragile.
- Increased volatility risk when conditions change rapidly or central bank decisions are unexpected.

Challenges for monetary policy communication

- AI-mediated interpretation may make it harder for central banks to reach human audiences effectively.
- If central banks also use AI to craft messages, machines may end up communicating with machines.
- Concerns about communication effectiveness and feedback loops that amplify market moves or create self-fulfilling expectations.
- Market participants and central banks should be aware of these implications.
- Critical examination of AI-assisted analysis is essential to manage risks and retain control over the impact of monetary policy communication.