

Global Database for Central Bank Communications

NLP Infrastructure for Monetary Policy Analysis

Cory Baird, Jonathan Benchimol, Vira Vyshnevska, Iegor Vyshnevskyi, and Wook Sohn

February 18, 2025

This presentation does not necessarily reflect the views of the Bank of Israel or any other institutions.

Outline

Introduction

GDCBC: Technical Infrastructure

Dissecting CB Statements

Generative AI Application

Summary & Future Work

Appendix

Not Just a Database: Reproducible AI for Macroeconomics

- ▶ **Scientific Rigor**

Infrastructure to maintain scientific standards in text analysis.

- ▶ **Unprecedented Scale**

Largest collection of **standardized** central bank (CB) communications.

- ▶ **Cutting-Edge Infrastructure**

Transparent, automated, and scalable pipelines.

- ▶ **Applying AI to Macroeconomics**

Leverages AI to identify economic narratives and sentiment over time and across countries.

- ▶ Today: Global Financial Cycle origins and drivers (Rey, 2015).

Current Literature

Key Limitations in CB Communication Research:

- ▶ **Data & Standardization:** Existing research focuses only on a few countries or uses only BIS speeches (Born et al., 2014; Picault and Renault, 2017).
 - ▶ Contribution: standardized dataset of 50 CBs
Larger than existing non-public databases, e.g., Gonzalez and Tadle (2022)
- ▶ **Reproducibility:** Lack of open-source pipelines impedes reproducible research (Gentzkow et al., 2019; Hansen and McMahon, 2016).
 - ▶ Contribution: transparent and accessible text cleaning, and Large Language Model (LLM) pipelines

Applying the Scientific Method to "AI" and NLP

- ▶ Failure to share data/algorithms hinders reproducibility and efficiency
 - ▶ **Can reproduce** lexical/dictionary methods at immense cost
 - ▶ **Cannot reproduce** black box LLMs (Claude, Gemini) because data, prompts, and models must be **exactly** the same
- ▶ **Irreproducible NLP methods** hinders scientific progress and data reliability in downstream modeling

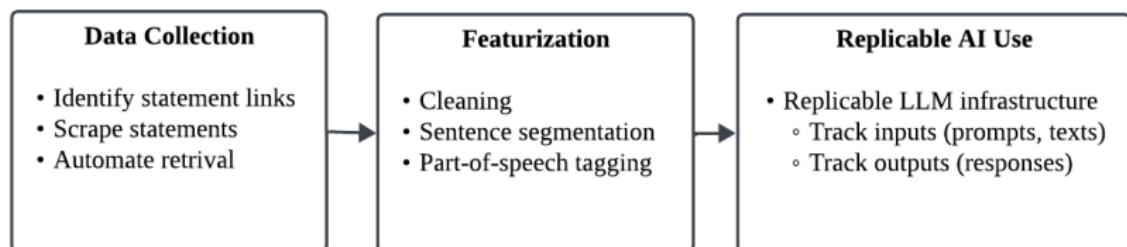


Figure 1: End-to-End Transparency for Replicable AI

Infrastructure Overview

- ▶ **Transparency**
 - ▶ All code/scripts made public on GitHub
 - ▶ Technologies: GitHub, Pyenv, Poetry
- ▶ **Automation**
 - ▶ Scripts executed regularly via a Data Version Control (DVC) pipeline
 - ▶ Technologies: DVC, GitHub Actions.
- ▶ **Scalability**
 - ▶ Modular scripts easily adaptable for other CB texts
 - ▶ Technologies: Python Object Oriented Programming, GitHub
- ▶ **Reproducibility**
 - ▶ Transparent text analysis
 - ▶ Reproducible pipelines for cleaning, featurization, and LLMs
 - ▶ Technologies: Pyenv, Poetry, S3, DVC, Weights & Biases

GDCBC Overview

- ▶ Monetary Policy Statements
 - ▶ Scope
50 CBs (35 API and 15 static currently cleaned)
 - ▶ Data
Approximately 6,066 unique MPSs
 - ▶ Temporal Coverage
1990-Present

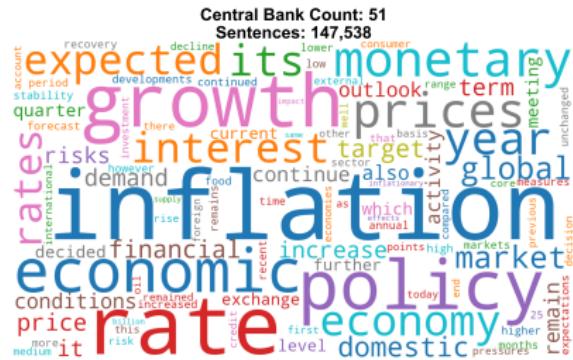


Figure 2: Standardized and diverse corpus for training/testing A.I.

Outline

Introduction

GDCBC: Technical Infrastructure

Dissecting CB Statements

Generative AI Application

Summary & Future Work

Appendix

Technical Infrastructure: Code & Data management

- ▶ Systematic and replicable data collection
 - ▶ Code version control: Git and GitHub
 - ▶ Python version control: pyenv and uv
 - ▶ Data version control: DVC
 - ▶ Pipeline management: DVC
 - ▶ Package/Library management: Poetry and uv
 - ▶ Storage: AWS S3
- ▶ LLM infrastructure
 - ▶ LLM experimentation: Weights and Biases
 - ▶ GPU compute: Google Colab

Code/Folder Infrastructure: Automated & Scalable

- ▶ **Modular Design:** Programs/scripts are modularized for:

1. Text retrieval
2. Text cleaning
3. LLM inference

- ▶ **Standardized Folder Structure:**

`src` : Python scripts

`data` : Text files

`references` : GenAI prompts

- ▶ **GDCBC Updates:** Automated updates using DVC.

- ▶ **End-to-End Reproducibility:** A full pipeline ensures reproducible results from data acquisition to NLP analysis.



Figure 3: Folder structure: "cookiecutter" template

Outline

Introduction

GDCBC: Technical Infrastructure

Dissecting CB Statements

Generative AI Application

Summary & Future Work

Appendix

The Rise of Central Bank Communication

- ▶ CB communication dramatically changed since the 1990s
 - ▶ First official MPS: Reserve Bank of Australia in 1990
 - ▶ Rapid growth in MPSs after the year 2000
- ▶ MPSs are a crucial tool for:
 - ▶ Managing inflation expectations
 - ▶ Enhancing policy predictability
 - ▶ Building credibility and trust

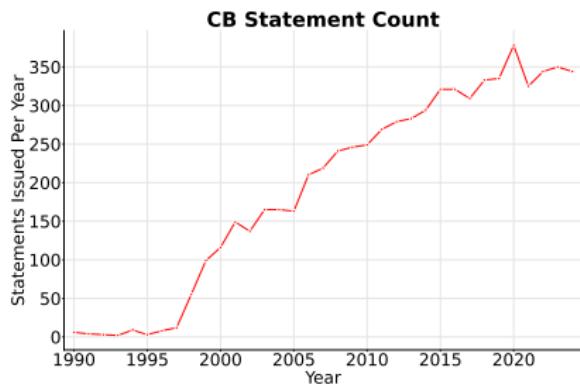


Figure 4: Annual MPSs in the GDCBC

Descriptive Table

Country		ISO2	First Statement	Last Statement*	Count (Total)	Count (2023)
Argentina		AR	2016-Aug-09	2024-Nov-01	55	4
Australia		AU	1990-Jan-23	2024-Dec-10	231	11
Brazil		BR	2006-Mar-08	2024-Dec-11	151	8
Canada		CA	2009-Jan-20	2024-Dec-11	128	8
Switzerland		CH	2000-Jan-20	2024-Dec-12	117	4
Chile		CL	1998-Jan-08	2024-Oct-17	287	8
Colombia		CO	2015-Aug-21	2024-Oct-31	85	9
Czechia		CZ	2014-Feb-06	2024-Nov-07	88	8
European Central Bank		ECB	1999-Mar-04	2024-Dec-12	304	8
Egypt		EG	2005-Jun-02	2024-Nov-21	121	0
Great Britian		GB	1997-Jul-16	2024-Nov-07	202	8
Ghana		GH	2011-Feb-18	2024-Nov-29	71	6
Gambia		GM	2015-Feb-24	2024-Nov-26	37	4
Hungary		HU	2002-Dec-16	2024-Nov-19	369	13
Indonesia		ID	2005-Aug-16	2024-Nov-20	181	12
Israel		IL	2002-Sep-23	2024-Nov-25	131	8
India		IN	2015-Apr-07	2024-Dec-06	59	6
Iceland		IS	2009-Jan-29	2024-Nov-20	121	6
Japan		JP	1998-Jan-16	2024-Dec-19	358	8
Kenya		KE	2015-Aug-05	2024-Dec-05	55	7
South Korea		KR	1999-Dec-02	2024-Nov-28	233	8
Kazakhstan		KZ	2006-Sep-05	2024-Nov-29	85	8
Sri Lanka		LK	2012-May-11	2024-Nov-27	122	8
Liberia		LR	2019-Nov-20	2024-Aug-27	18	3
Lesotho		LS	2017-Apr-04	2024-Nov-26	43	6

Country		ISO2	First Statement	Last Statement*	Count (Total)	Count (2023)
Moldova		MD	2001-Jan-01	2024-Dec-05	202	8
Mongolia		MN	2010-Jul-23	2024-Dec-13	60	4
Malawi		MW	2014-Apr-30	2024-Nov-04	33	4
Mexico		MX	2018-May-17	2024-Dec-19	55	8
Malaysia		MY	1997-Oct-18	2024-Nov-06	140	6
Mozambique		MZ	2017-Oct-26	2024-Nov-27	44	6
Nigeria		NG	2004-Feb-27	2024-Nov-26	183	9
Norway		NO	1996-Nov-05	2024-Dec-19	210	8
New Zealand		NZ	1999-Mar-17	2024-Nov-27	171	7
Philippines		PH	2001-Dec-26	2024-Dec-19	218	9
Pakistan		PK	2005-Dec-01	2024-Dec-16	82	9
Poland		PL	2001-Nov-28	2024-Dec-04	248	11
Rwanda		RW	2019-Feb-07	2024-Nov-20	11	0
Sweden		SE	1996-Jan-09	2024-Dec-19	191	5
Sierra Leone		SL	2015-Dec-14	2024-Dec-23	18	5
Eswatini		SZ	2011-May-13	2024-Nov-22	59	6
Thailand		TH	2000-May-23	2024-Dec-18	194	6
Tajikistan		TJ	2017-Jan-31	2024-Apr-26	22	4
Turkiye		TR	2005-Jan-03	2024-Dec-26	223	12
Taiwan		TW	2001-Jun-28	2024-Sep-19	98	4
Ukraine		UA	2014-Apr-15	2024-Dec-13	87	8
Uganda		UG	2011-Jul-01	2024-Dec-05	92	6
United States		US	1994-Feb-04	2024-Dec-18	234	8
Uzbekistan		UZ	2020-Jan-18	2024-Dec-12	21	8
South Africa		ZA	1999-Oct-13	2024-Nov-21	161	6
Zambia		ZM	2016-Nov-16	2024-Nov-13	31	4

Figure 5: MPS Information

Dataset Coverage

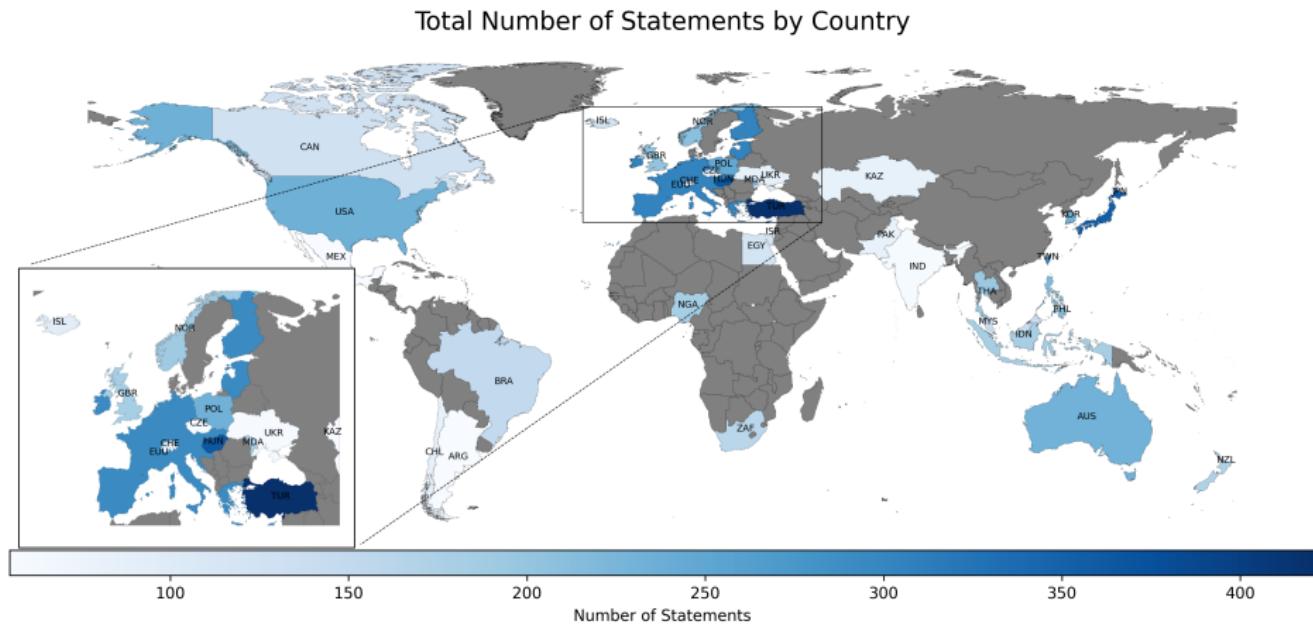


Figure 6: Countries and Monetary Policy Statements

Dataset Coverage

- ▶ The MNB (Hungarian National Bank), BoJ, and ECB have issued the most statements
- ▶ CBs in middle-income countries tend to release fewer statements and have shorter publication histories

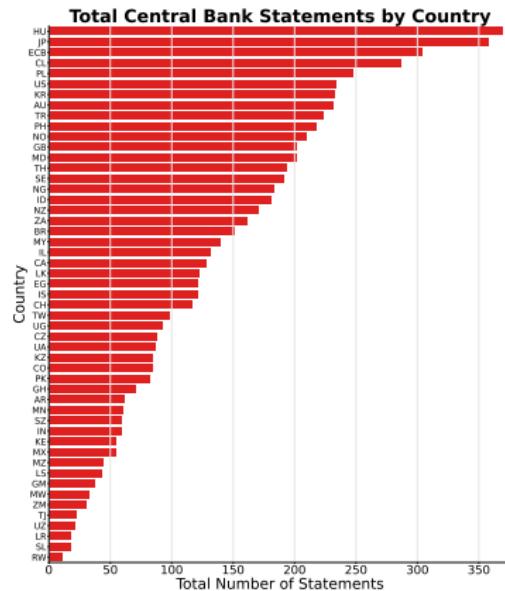


Figure 7: Statements by Country

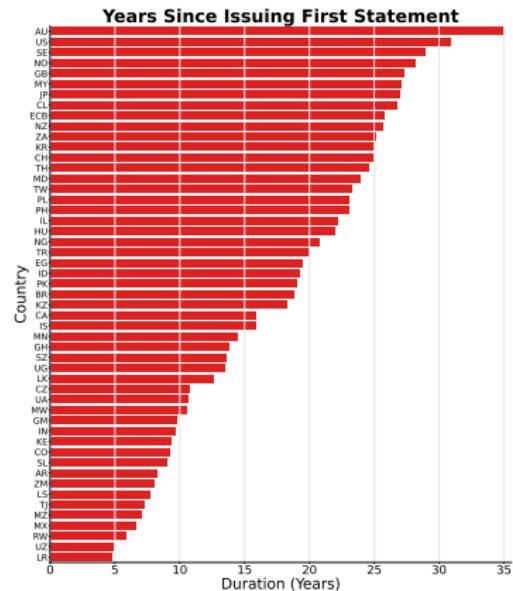


Figure 8: Years of Coverage by Country

Trends in Communication Over Time

- ▶ MPSs growing lengthier over time (proxied by average adjective per statement)
- ▶ MPSs becoming slightly more "readable" (measured as the average of multiple readability indexes, such as the Coleman-Liau Index, which assign US grade-level equivalents to text statements)

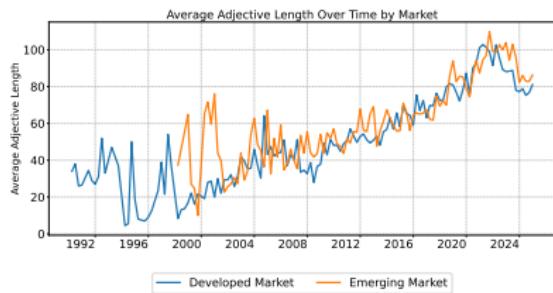


Figure 9: MPS Length
(avg. adjectives per MPS)

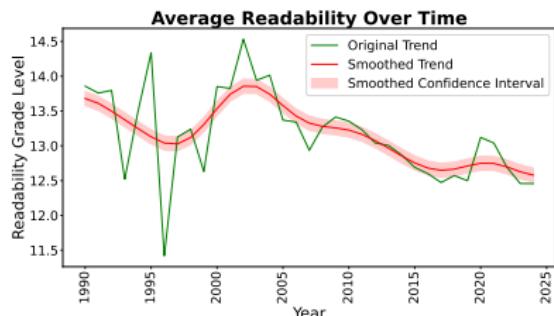


Figure 10: Readability Over Time

Appendix: Visualizing Monetary Policy Keywords Over Time

Key Characteristics (Patterns) of the MPS

- ▶ Typical structure includes:
 - ▶ Economic forecasts
 - ▶ Economic analysis (Domestic & Foreign)
 - ▶ Policy (Interest Rates & Forward Guidance)
 - ▶ Comments on Policy Mix (Fiscal, Trade)

Econ. Theme	Part-of-Speech (POS)	Example	Spacy part-of-speech Matcher
Economic Forecasts	NOUN+VERB+NUM	GDP to grow 3%	POS: NOUN, POS: VERB, POS: NUM
Economic Forecasts	NOUN+VERB+ADJ+NOUN	Markets expect stronger growth	POS: NOUN, POS: VERB, POS: ADJ, POS: NOUN
Economic Analysis (Domestic)	NOUN+PREP+NOUN	Risk of recession	POS: NOUN, POS: PREP, POS: NOUN
Economic Analysis (Foreign)	NOUN+PREP+NOUN	Impact of Brexit	POS: NOUN, POS: PREP, POS: NOUN
Policy (Interest Rates)	NOUN+VERB+NUM	Rates to rise 0.5%	POS: NOUN, POS: VERB, POS: NUM
Policy (Forward Guidance)	ADJ+NOUN+VERB	Tighter policy expected	POS: ADJ, POS: NOUN, POS: VERB
Comments on Policy Mix (Fiscal)	NOUN+VERB+NOUN	Budget to support growth	POS: NOUN, POS: VERB, POS: NOUN
Comments on Policy Mix (Trade)	NOUN+PREP+NOUN	Impact of tariffs	POS: NOUN, POS: PREP, POS: NOUN

Table 1: Examples of MPS Sentence Patterns by Part-of-Speech

Most Common Sentence Patterns by Part-of-speech

- ▶ Common sentence patterns include:
 - ▶ DET ADJ NOUN (e.g., "The strong dollar," "The current rate")
 - ▶ DET NOUN ADP (e.g., "The impact of," "The level of")
 - ▶ ADP DET NOUN (e.g., "In the market," "On the economy")
 - ▶ Sentences identified with spaCy's Universal POS tags and pattern matching.

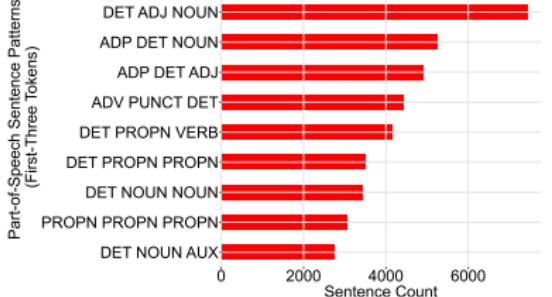


Figure 11: Most Common Patterns by Part-of-Speech (POS)



Figure 12: POS: The DET PROPN VERB pattern is associated with CB commentary/actions

Outline

Introduction

GDCBC: Technical Infrastructure

Dissecting CB Statements

Generative AI Application

Summary & Future Work

Appendix

Addressing AI Challenges with the Corpus/Infrastructure

► Challenges with LLMs

1. Irreproducibility: Insufficient transparency regarding model inputs and specifications makes replicating it difficult.
 - ▶ We provide open-source code, including LLM code, prompts, and Jupyter notebooks, ensuring full transparency and reproducibility.
2. High computational costs and energy consumption.
 - ▶ We implement text filtering (keyword & part-of-speech) pipelines to reduce the computational burden.

► Application to Econ: The Global Financial Cycle (GFC)

- ▶ Communication Channel: CB Communication ⇒ Market Expectations ⇒ Financial Conditions
- ▶ Using our LLM/Data Pipeline we leverage LLMs for Question Answering (QA) & classification (topic, sentiment) tasks to explore origins and drivers of the Global Financial Cycle (Miranda-Agrippino and Rey, 2020, 2022)

Leveraging Our LLM/Data Pipeline

- Infrastructure is abstract for scalability
 - LLM: seamlessly add more models (ChatGPT, Huggingface, etc) and tasks (classification, summarization, Q&A)
 - Text retrieval: seamlessly add more central banks and types of communications (minutes, speeches, etc)

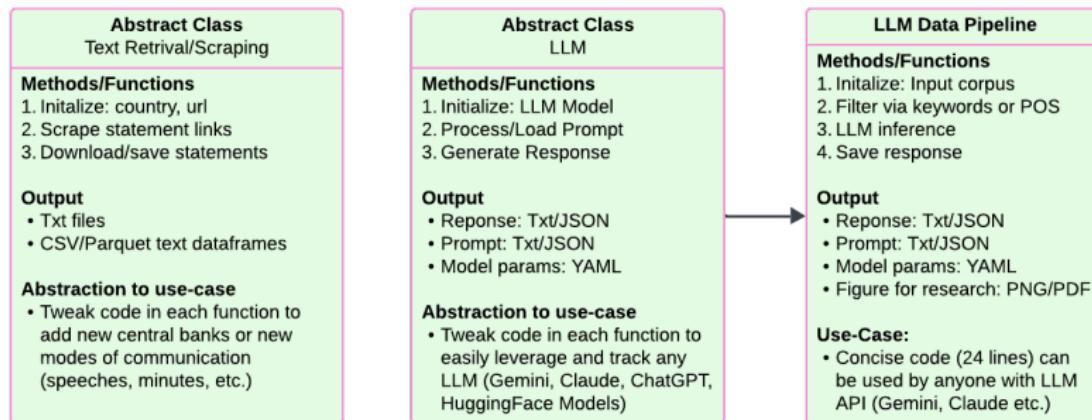


Figure 13: Comparing Text Retrieval and LLM scripts/programs

LLM Application: Exploring the Global Financial Cycle

- ▶ LLM Application: Question Answering (QA)
 - ▶ Prompts model to summarize drivers of global economy

"I want you to identify the main driver of global economic and financial conditions and the origin country driving these conditions. In the summary please note if any central banks identify the Federal Reserve as the driver of global economic or financial conditions."
- ▶ LLM Application: Aspect Based Sentiment¹(ABS)
 - ▶ Prompts model to extract sentiment for specific themes

"Analyze the provided central bank text and identify any discussions related to the impact of global financial conditions on domestic economies, focusing on themes such as economic growth, financial conditions, capital flows, monetary policy spillovers. Determine the sentiment (Positive, Negative, or Neutral) expressed towards each identified theme. Also identify the related country."
- ▶ Findings from LLM Analysis:
 - ▶ QA: Central Banks often cite the U.S. economy as a driver of global economic conditions rather than specifically the U.S. Federal Reserve.
 - ▶ ABS: Sentiment analysis shows that central banks expressed more concern about growth and financial conditions rather than monetary policy and capital flows in the post-2008 period

Appendix: Full Prompt Example

¹Aspect-based sentiment identifies sentiment of specific theme

LLM Application: Summarization (QA)

Comparing LLMs

Gemini

Mechanism: Global economic conditions
Origin Country: Global Summary: The central bank comments indicate that the main driver of global economic conditions is the ongoing global economic slowdown and deterioration, with concerns about the stability of the global financial system and the pace of global economic recovery. Many countries are experiencing a softening of domestic economic growth due to weaker external demand and the adverse effects of the global economic turmoil. The central banks are closely monitoring global financial and economic developments and taking necessary measures to mitigate the impact on their respective domestic economies.

Claude

Mechanism: Global economic slowdown/crisis, particularly in major economies, impacting global credit markets and investor confidence. This translates to a decline in global demand, impacting financial stability and wealth. Origin Country: United States (implied, given the context of "major economies"). While multiple countries mention "global" economic problems, the specifics in the analysis suggest an origin in the financial turmoil sparked by issues in US (and possibly other western) financial markets. Summary: Central banks worldwide report that the global economic slowdown, originating in significant financial market problems in major economies, is the primary driver of economic weakness across the globe. Weakened confidence and reduced demand are creating widespread economic deterioration. Recovery is seen as dependent on restoring financial stability globally.

Figure 14: Summary of the same text differ by LLM

LLM Application: Aspect Based Sentiment

Sentiment Based Aspect: LLM Claude

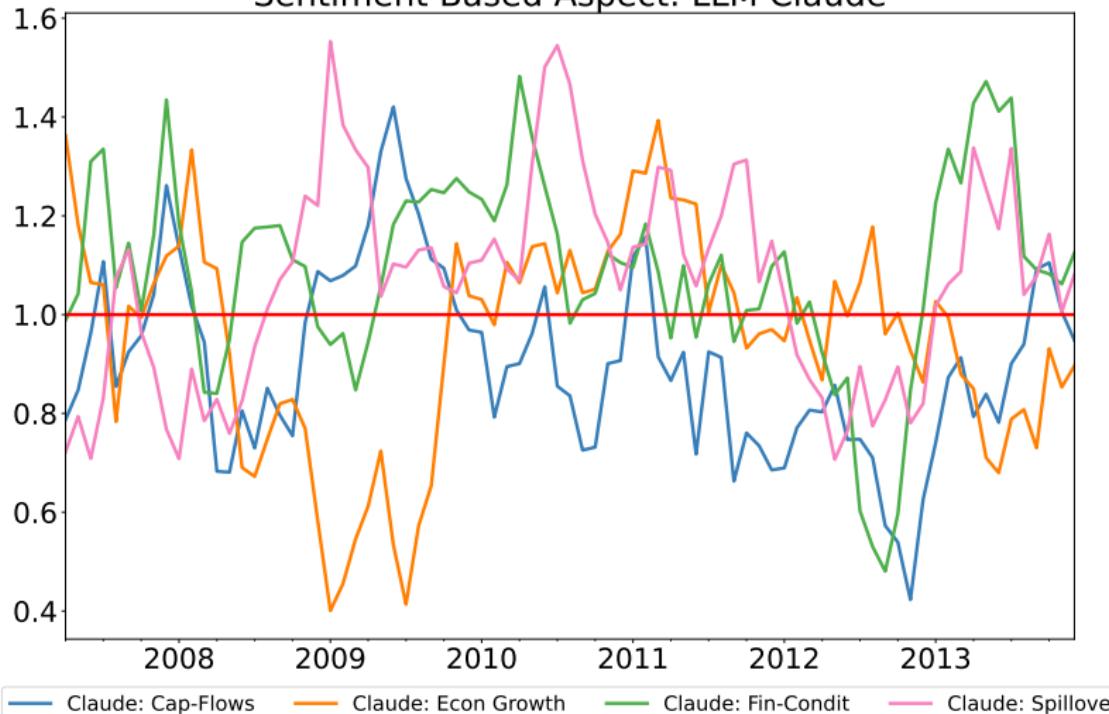


Figure 15: LLM can easily produce aspect base sentiment

Sentiment: Lexical Outperforming LLM

Global Sentiment vs. Global Financial Cycle

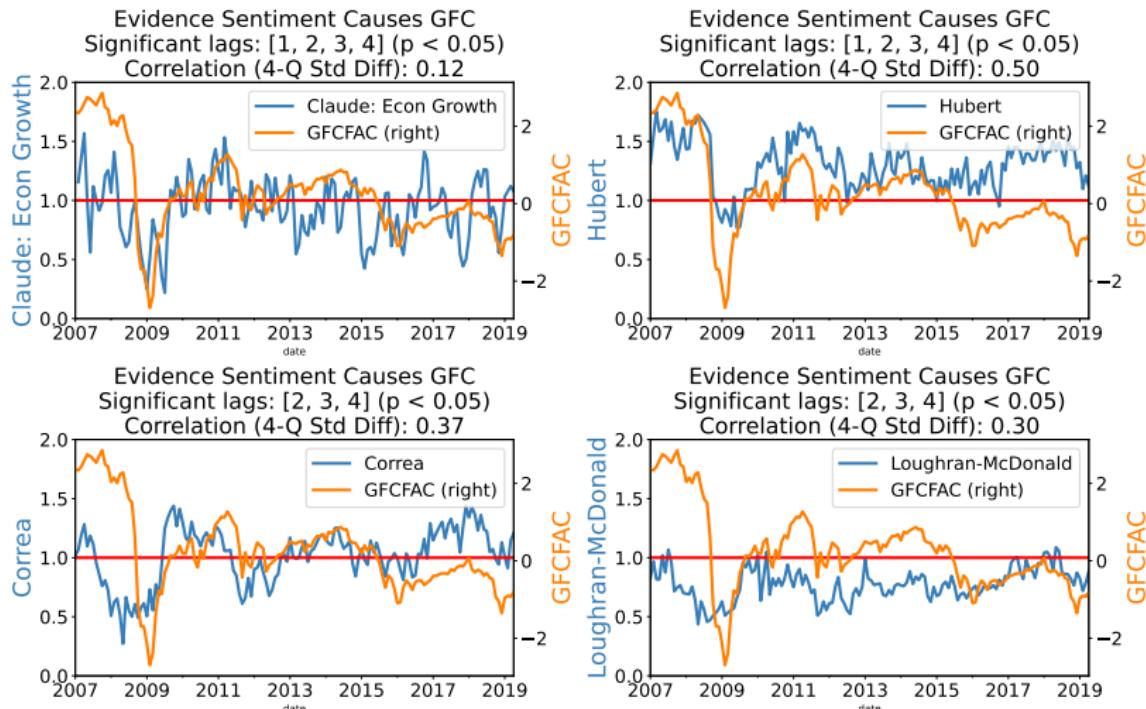


Figure 16: LLM produces aspect base sentiment indices

Evidence that Sentiment Drives GFC

- ▶ High correlation between the 4Q difference of Hubert sentiment and the Global Factor
- ▶ Granger tests show that DM and EM aggregated sentiment (avg.) Granger causes the GFC (4Q-difference)
- ▶ GFC does not Granger causes sentiment changes

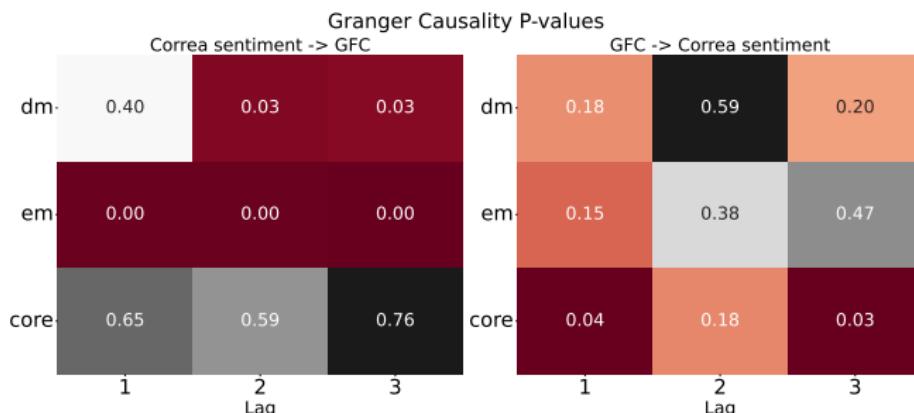


Figure 17: Hubert Sentiment Granger Causes GFC

Core: ECB, AU, CA, CH, GB, JP, and US

Outline

Introduction

GDCBC: Technical Infrastructure

Dissecting CB Statements

Generative AI Application

Summary & Future Work

Appendix

Summary & Future Work

- ▶ We construct an open-source standardized database of CB communications designed as a benchmark for LLM tasks.
- ▶ We distribute all source code for our Data/LLM infrastructure to allow researchers to leverage AI tools in a scalable and reproducible manner.
- ▶ Using LLMs, we find that central banks often cite the U.S. economy broadly (rather than mention the U.S. Federal Reserve or monetary policy spillovers) as the primary driver of global conditions.
- ▶ We plan to scale our data to include more central banks and communication types.
- ▶ We will include most major closed-source (Gemini, Claude, ChatGPT) and open-source (Hugging Face) LLM systems in our easy-to-use infrastructure.

Outline

Introduction

GDCBC: Technical Infrastructure

Dissecting CB Statements

Generative AI Application

Summary & Future Work

Appendix

Monetary Policy Statements

The screenshot shows the Bank of Canada's website with a search bar at the top. Below it, there are links for Head, Monetary Policy, Financial Stability, Statistics, Banknotes and Coins, Publications and Research, and a search icon. The main content area is titled "Copom Statements". It features a sidebar with a dropdown menu for "Copom" and a list of statements grouped by month: November (1), September (1), August (1), and June (1). Each statement entry includes a brief description of the policy change.

Month	Number of Statements	Description
November	1	Copom reduces the Selic rate to 12.25% p.a.
September	1	Copom reduces the Selic rate to 12.75% p.a.
August	1	Copom reduces the Selic rate to 13.25% p.a.
June	1	Copom maintains the Selic rate at 13.75% p.a.

(a) Brazil: Statements in press section

The screenshot shows the Bank of Canada's website with a search bar at the top. Below it, there are links for ABOUT US, CORE FUNCTIONS, MARKETS, BANKNOTES, PUBLICATIONS, RESEARCH, PRESS, and STATISTICS. The main content area is titled "Press releases". It features a sidebar with filters for Contains (Keywords), Locations (All Locations), Sources (All Sources), Topics (All Topics), and Published After (1999-09-01). The results count is 785 result(s). The first result is a news item about Nick Leswick's appointment as Executive Director of Policy.

785 result(s)

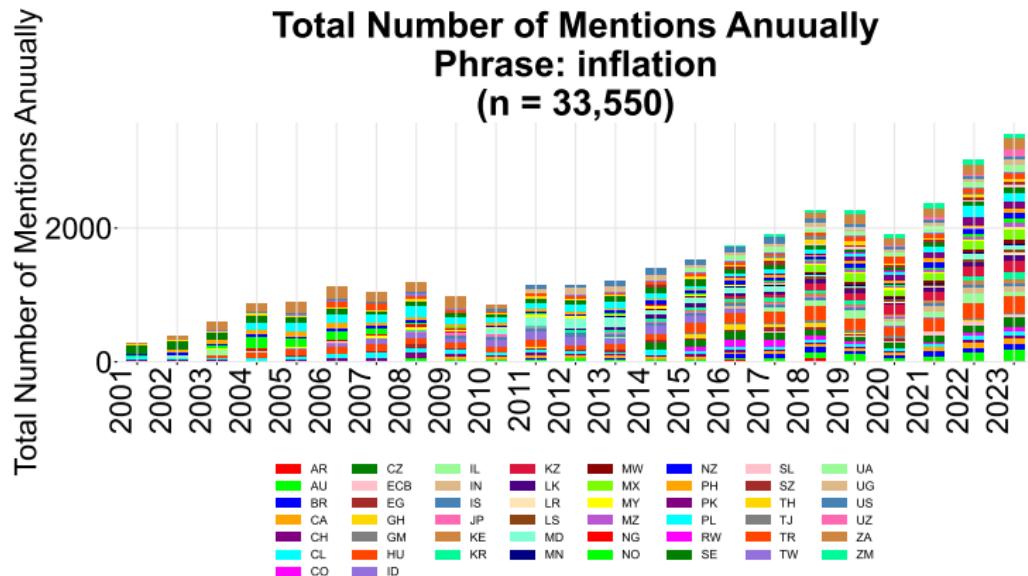
Bank of Canada announces appointment of Nick Leswick as Executive Director of Policy

Media Relations - Ottawa, Ontario
The Bank of Canada today announced that Nick Leswick has been appointed to the new role of Executive Director of Policy, effective January 3, 2023.
Current Topic: Press, Press releases

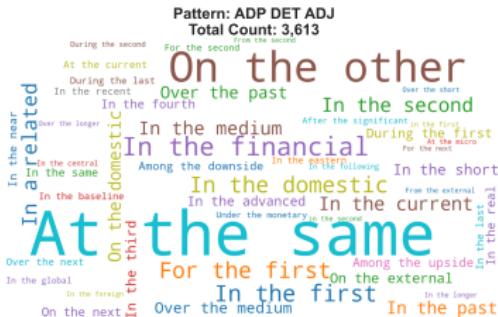
(b) Canada: Statements in press section

Figure 18: Examples of MPSs hosted on central bank websites

Key Words Time Series



Appendix: Part-of-Speech Word Clouds



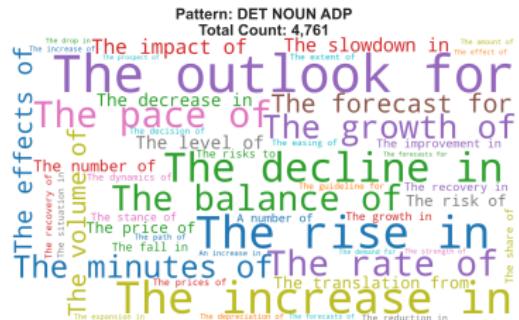
(a) ADP DET ADJ



(c) DET ADJ NOUN



(b) ADP DET NOUN



(d) DET NOUN ADP

Figure 20: Common Part-of-Speech Patterns at the Start of Sentences

Leveraging AI to Extract Economic Narratives

- ▶ Leveraging Our LLM/Data Pipeline
 - ▶ We leverage LLMs for Question Answering (QA) & classification (topic, sentiment) tasks to explore origins and drivers of the Global Financial Cycle (Rey, 2015)
 - ▶ Using our pipeline to extract narrative we:
 1. Search keywords (global, spillover, flows): 6,135 sentences
 2. Truncate for 2019 data only: 1,228 sentences
 3. Randomly select one sentence for each country: 24 sentences

Prompt & Input Text

The following texts comes from central banks. The country is identified in the Country: and the Comment: is the text released by the central bank. I want you to identify the main driver of global econ ...

Text to analyze:
Country: AU, Comment: With considerable economic policy stimulus in train around the world, the global economy is resuming growth. Sentiment in global financial markets has continued to improve. This ...

Country: CA, Comment: While significant fragilities remain, global economic developments have been slightly more positive and the global outlook has improved modestly relative to the Bank's projection....

Country: HU, Comment: Over recent months, global economic conditions and the prospects for growth both have deteriorated more sharply than previously expected.

Country: ID, Comment: The improvement in the global economic outlook and developments has met with positive response visible in various financial indicators, including world stock indices, and the sha....

Country: IS, Comment: Considerable uncertainty remains concerning external debt, government financing and financial sector restructuring, and the global environment remains difficult.

Country: ML, Comment: At the Monetary Policy Committee meeting today, Bank Negara Malaysia decided to leave the Overnight Policy Rate (OPR) unchanged at 2.00 %. Since the previous MPC meeting, signs of....

Country: MY, Comment: At the Monetary Policy Committee meeting today, Bank Negara Malaysia decided to leave the Overnight Policy Rate (OPR) unchanged at 2.00 %. Since the previous MPC meeting, signs of....

Country: NG, Comment: While this development augurs well for Nigeria's fiscal and external sector positions , its sustainability would, however, depend on how quickly the global economy would bottom

Country: NO, Comment : The outlook for the global economy has deteriorated.

Country: NZ, Comment: As a result, New Zealand has not benefited to any significant extent from the rebound that has occurred recently in global hard commodity prices.

Country: PH, Comment: These changes would provide more liquidity and credit in the banking system to ensure the orderly functioning of financial markets should global financial conditions worsen. The

Country: TW, Comment: Global economic outlook continues to worsen. The global financial crisis triggered by the U.S. housing bust has severely depressed the world economy. A relatively high degree of

Country: US, Comment: Furthermore, global demand appears to be slowing significantly.

Country: ZA, Comment: With respect to economic growth, the domestic economy is being adversely affected by the continuing turbulence in the global economy. The outlook for oil prices remains uncertain....

Please respond in the following format:

Text analysis: [Identified Issue/Topic]

Origin Country: [Origin Country]

Summary: [Brief explanation]

LLM Response

Global financial crisis triggered by the US housing market bust, resulting in slowing global demand and depressed global economy.

origin country*****

US

Central bank comments across a variety of countries overwhelmingly point to the US housing market collapse as the initial trigger for a global financial crisis.

This crisis is causing significantly depressed global growth, reduced global demand, and considerable volatility in global financial markets.

The consequences are being felt worldwide, negatively impacting various economies, as evidenced by concerns about exports, oil prices, and stock markets.

Regional Sentiment vs. Global Factor

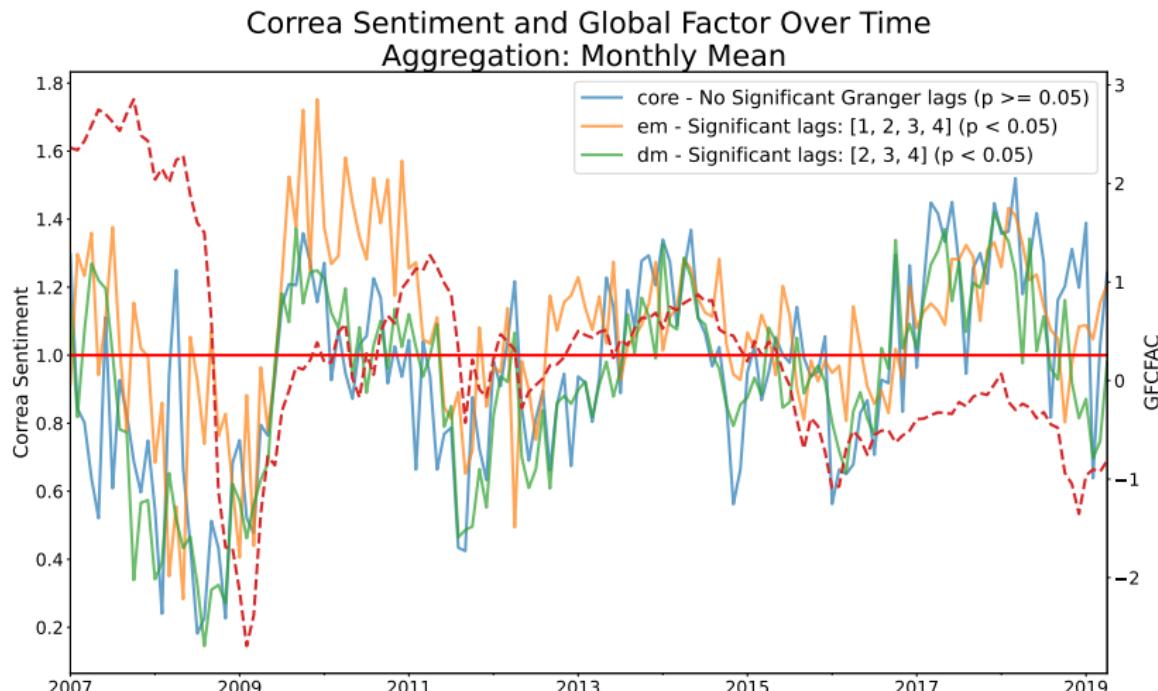


Figure 21: Regional Sentiment

References I

- Born, B., Ehrmann, M., and Fratzscher, M. (2014). Central bank communication on financial stability. *Economic Journal*, 124(577):701–734.
- Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–574.
- Gonzalez, M. and Tadle, R. C. (2022). Monetary policy press releases: an international comparison. BIS Working Papers 1023, Bank for International Settlements.
- Hansen, S. and McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99(S1):114–133.
- Miranda-Agricoppino, S. and Rey, H. (2020). U.S. monetary policy and the Global Financial Cycle. *Review of Economic Studies*, 87(6):2754–2776.

References II

- Miranda-Agrrippino, S. and Rey, H. (2022). The global financial cycle. In *Handbook of International Economics*, volume 6, pages 1–43. Elsevier.
- Picault, M. and Renault, T. (2017). Words are not all created equal: A new measure of ECB communication. *Journal of International Money and Finance*, 79:136–156.
- Rey, H. (2015). Dilemma not trilemma: the global financial cycle and monetary policy independence. NBER Working Papers 21162, National Bureau of Economic Research, Inc.