Mapping Latent Narrative Content Using LLMs: Essay on Economic Uncertainty Drivers

Sami Diaf[1,*,†], Florian Schütze[2,†]

[1]Universität Hamburg, Department of Socioeconomics, Hamburg
[2]Helmut-Schmidt-Universität, Faculty of Economics and Social Sciences, Hamburg

Abstract

The use of Large Language Models (LLMs) has been widespread, and their adoption by practitioners opened ways to deepen the analyses at unprecedented granular levels. This facilitated implementing text-as-data applications in social sciences, which previously required substantial time and effort, for instance in narrative economics. Investigating uncertainty has been of key importance in modern mainstream economics, especially at central banks, where it is considered as a pillar of monetary policy practice. This work sets an LLM-based framework to reveal the narratives of uncertainty drivers in selected monetary policy documents, yielding a Directed Acyclic Graph (DAG) structure featuring causal patterns and their intensities. Besides its topic-rich, highly informative network structure, results were found to meet event-based documents and the prevailing macroeconomic context. Although LLMs are known to be bias-sensitive and difficult for replications, establishing a central node in the output graph enabled a powerful narrative map in the Federal Reserve governors speeches without requiring few-shot strategies or metadata incorporation. Tested on two minutes issued during the 2008 financial crisis and the COVID-19 pandemic, it successfully retrieved latent factors pertaining to monetary policy uncertainty and robust context-aware narrative quality thanks to LLMs reasoning abilities. Moreover, this scheme outperforms deep semantic search models trained on a big central bank corpus and has the advantage of revealing more granular patterns for specific context-based tests.

Keywords

LLMs, narrative economics, DAG, uncertainty

1. Introduction

Innovations in the field of natural language processing (NLP) played a key role in maximizing information extraction from text documents and enhancing the performance of several tasks whose execution requires a substantial semantic depth.

The advent of LLMs, as deep learning models trained on "next-word" prediction tasks, dramatically fostered information extraction and eased the application of text-as-data methods. As neural models, LLMs feature memory-based architectures called transformers [1], able to capture long-range dependencies in training data.

Although hailed by users for their high accuracy and speed of execution, LLMs remain black-box models whose inner architecture was trained on various documents, making the output prone to judgment errors or interpretation. Their use is not bias-free and is constrained regarding the hyperparameter tuning and the data quality [2], and they are considered by many as *stochastic parrots* [3].

In empirical research, the use of LLMs has been widespread but remains a big challenge. Practitioners interact with LLMs using prompts and do ignore the documents used during the training step, while the output may not be unique, and its variability fuels reproducibility issues and biases [3]. Recent advances shifted the use toward a multitasking strategy using AI agents to better handle the LLMs

ESWC'26: 5th Workshop on LLM-Integrated Knowledge Graph Generation from Text (TEXT2KG), May 10-14, 2026 | Dubrovnik, Croatia.
*Corresponding author.
†These authors contributed equally.
sami.diaf@uni-hamburg.de (S. Diaf); schuetze@hsu.hamburg (F. Schütze)
0000-0002-9554-3475 (S. Diaf); 0009-0002-5512-0547 (F. Schütze)

output. The *bona fide* of LLMs was demonstrated in political science [4], where few-shot promptings outperformed automated content analysis.

Parallel to the development of NLP applications, narrative economics [5] established itself as a leading research field, particularly in monetary policy. The literature witnessed an unprecedented interest of researchers trying to uncover textual information for testing and inference purposes [6], usually with architectures describing the narrative statements via imbricated NLP methods [7].

Uncertainty plays a pivotal role in the monetary policy practice and is considered a forward guidance tool at the hands of central bankers [8]. First attempts to gauge uncertainty in text documents used basic word count methods, as for the economic policy uncertainty [9]. Some authors [10] employed a semantic search approach [11] to better apprehend contextual information and distinguish uncertainty from risk, based on a large corpus of collected central bank speeches. Moreover, the notion of uncertainty within an economic framework has been the subject of extensive academic scrutiny [12].

While semantic search models require a sizable amount of documents to achieve robust results, they do offer better results than dictionaries but at a questionable, often agnostic topic quality [13] and an extensive training time. Further hyperparameter selection and post-hoc adjustments [14] are usually used to enhance the topic structure, requiring significant resources. LLMs are considered deep embedding models, deriving their power from their flexible context-based information, in contrast to traditional static word embeddings [15, 16] and semantic search models [11]. Instead of relying on autoregressive schemes for token generation, LLMs are capable of *thinking*, which means generating a sequence of tokens representing intermediate steps in the reasoning process [17].

Recent attempts to use LLMs in monetary policy targeted stock market dynamics [6], emotions [18] and inflation expectations [19], to name a few. Results suggested a cautious use of LLMs for complex tasks requiring an in-depth analysis [18], and warned over the error-prone, customized prompting. These shortcomings stem from the often ignored behavioral patterns and cognitive processes [20].

This work tries to reveal uncertainty factors and their causal linkage in selected corpora related to monetary policy. It leverages the reasoning power of LLMs to track uncertainty formation and its lineage. Tested on two selected Federal Reserve minutes, this methodology details uncertainty drivers and the impact of macroeconomic factors on policymaking. Noticeable minutes from the 2008 financial turmoil and the 2020 pandemic demonstrated the importance of context information in discerning factors from events. Such granular patterns form a storytelling that cannot be retrieved using advanced topic models but does require human-like reasoning prone to a biased judgment.

Besides the high-quality content of uncertainty mapping, the application did not encounter hallucination claims even when setting higher values of the temperature parameter for deep information retrieval. This allows further policy-related tests to be conducted in terms of uncertainty resilience and risk management related to the retrieved drivers.

Findings recommend the use of zero-shot prompting for document scaling in comparative politics, as uncovered uncertainty factors were found to explain the rhetoric far better than automated text analysis. Moreover, LLMs offer a substantial narrative quality by remaining robust to policy changes and dynamic corpora, able to discern conventional from unconventional policy actions. The use of prior information, as for few-shot prompting, may enhance the results but comes with some risks if the used examples do not fully align with the scope of the corpus.

Although the experiment is prone to reproducibility issues linked to the choice of LLMs and their reasoning abilities, a multistep approach (Chain-of-Thought [21]) for uncertainty could enhance results, for example, at the level of economic agents or with regard to specific variables. Such schemes render more stable results, but at an expensive cost that hinders reproducibility and generalization.

The paper outlines the development of text analysis (Section 1) and then details uncertainty in text mining and the recent adoption of LLMs as a technology (Sections 2 and 3). Section 4 features an application on two distinct minutes and details the graph-like results, later discussed in Section 5.

2. Uncertainty in text analysis

Uncertainty is generally known as the state of being unsure or facing an unclear situation. It has been considered as a hidden variable that cannot be captured via simple, straightforward measurement methodologies.

In macroeconomics, uncertainty significantly reduces the effectiveness of devised policies. Mainstream economics experienced during long periods a fierce debate on how to define uncertainty and how to separate it from risks. Assessing sources of uncertainty remains a debated philosophical subject [22], whose origin stems from Keynes [23] and Knight [24]. The former Chairman of the Federal Reserve, Ben Bernanke [25], proposed three broad categories to separate uncertainty from risk: uncertainty about the state of the economy, the structure of the economy, and expectations. This classification pertains to distinguishing backward-looking variables (state and structure) from forward-looking ones (expectation).

Earlier attempts to gauge uncertainty in text mining used keywords [26] for a general-level estimation. This text-as-data strategy assessed policy uncertainty using the occurrence of specific terms in selected corpora [9, 27, 28], without providing context information on uncertainty colors [29].

While researchers still use these indices for investigative purposes, they lack semantic depth because uncertainty concerns, or alternatively confidence, could be mentioned by terms or expressions not encompassed by wordlists. Practitioners in economics are generally more concerned with uncertainty drivers and their amplitude than with uncertainty as a phenomenon. Several works attempted to explain uncertainty color [30] but failed to come up with concrete results, despite extensive efforts put on post-hoc inferences. This remains acute with the adoption of unconventional tools to steer monetary policy, whose implementation is subjected to policy changes. In other terms, uncertainty drivers could be ranked, depending on their outlook, into backward-looking and forward-looking drivers [29].

Semantic search models do leverage contextual information and provide a better assessment of high-level hypotheses compared to automated content analysis. [10] used semantic search models to unfold uncertainty origin and quantify its magnitude within the corpus of international central banks' speeches. The resulted embeddings confirmed a *knightian* approach that distinguishes uncertainty from risk adopted by leading central banks around the world. The methodology, built around *top2vec* [11] as an embedding topic model, could be considered as a shallow language model featuring a robust distributional representation given by the triplet document-topic-word, entirely depending on the corpus and on pre-trained models.

Solutions based on word embeddings [15] offer a granular analysis based on contextual patterns and require large corpora to achieve their results. However, they fail to come up with structured information that replaces human judgment or expertise.

LLMs are considered as *contextualized* word embeddings, acting as search engines and retrieving information from inputs they received during the training stage. Their reasoning capabilities are reliable alternatives to experts' assessments and can be used to yield structured output, tailored to the inquiries users request.

These advantages put LLMs above all previous text analysis techniques in terms of reliability, accuracy, and execution. Moreover, its dense architecture is capable of retrieving hidden factors related to a given task without requiring prior information or post-hoc inference. These characteristics fulfill the conditions to test uncertainty in documents, uncover its origin, and detail its drivers without designing sophisticated and expensive models.

3. Large Language Models in Economics

LLMs are deep neural networks featuring attention layers [1], serving as a memory mechanism to foster learning capabilities of the architecture. They act as search engines for information retrieval or powerful machines for question-and-answer tasks. Their capabilities have been praised and proven to outperform traditional automated content analysis and other NLP solutions [4], while remaining flexible and low-friction tools in political discourse analysis [31].

In economics, numerous applications have been given by researchers [19], praising *Generative AI* as reliable alternatives for economists [32], but mostly used for *shallow* tasks such as sentiment analysis.

Particularly, researchers in monetary economics used LLMs to solve issues related to specific classification and sentiment analysis. Earlier works targeted central bank communication, namely *CentralBankRoBERTa* [33] and *ECBERT-base* [34]. The latter is a deep learning model developed to detect emotions embedded in Federal Open Market Committee (FOMC) press conferences, while the former is a dual classifier for economic agents and sentiment analysis.

[29] used fine-tuned LLMs on monetary corpora in Japan to capture uncertainty dynamics. They built a monetary policy index based on forward-looking and backward-looking (respectively, anticipation-based and historical data) considerations on policy frameworks and market operations.

[35] conducted a randomness-type test on different uncertainty measurements and found LLMs successfully substitute panel experts for the case of Poland and the United States.

4. Experiment

An *OpenAI* model (*gpt-oss-120b*) [36] was used to uncover uncertainty factors in selected minutes released by the Federal Reserve, the organism steering monetary policy in the United States. The *gpt-oss-120b* model was designed for powerful reasoning, agentic tasks, and versatile developer use cases. Other LLMs, as for *meta-llama-3.1-8b-instruct*, failed to yield structured outputs and lacked a coherent lecture of uncertainty factors.

The experiment followed a zero-shot prompt strategy[1]. It did not provide any example or label to avoid guidance that may mislead the results. Over the last 25 years, Federal Reserve minutes dealt with various topics, such as banking stability, financial innovation, and crises, whose importance was fluctuating over time. Few-shot strategies might provide good anchors for uncertainty factors but would also distort results if the prompting overlooks debated topics.

Prompting (uncertainty factors)

system prompt = 'You are a macroeconomic expert in estimating uncertainty in a given text.
Evaluate the uncertainty drivers using a network, where each driver is a node and each vertex is the correlation between nodes.
The central node of the graph is "*monetary policy uncertainty*".
Final output are two dataframes: one for edges and one for vertices.'

Instead, we just required having *monetary policy uncertainty* as a central node in the graph-like generated answer. By doing so, we reinforce the goal of our task and provide a solid anchor as a starting point and test the output against event-based uncertainty.

For this aim, two minutes released during the 2008 financial crisis and the 2020 pandemic were selected to run the experiment of extracting uncertainty factors and check the retrieved nodes and their links.

At the height of the financial crisis, on September 16, 2008, the Federal Reserve released a statement geared toward stressed economic agents and financial markets. Figure 1 shows the output as a network structure centered around the node *monetary policy uncertainty*. It elicits 11 factors explaining uncertainty concerns, whose interactions offer a granular causality map. The reasoning ability of the *OpenAI* model helped derive a topic-like explanation of each node, as tabulated in Table 1.

Uncertainty sources from the financial sector form a distinct clique on the left side of the graph (*banking system stress*, *financial budget stress*, and *credit conditions tightness*), having the highest intensity links to *monetary policy uncertainty*. Intensity edges differentiate the factors' importance of financial and banking risks as primary uncertainty sources if compared to other sources (inflation and foreign exchange uncertainties are represented by two nodes each, at the bottom of Figure 1).

[1] maxTokens = 3000, temperature = 1.3, topP = 0.8, topK = 40

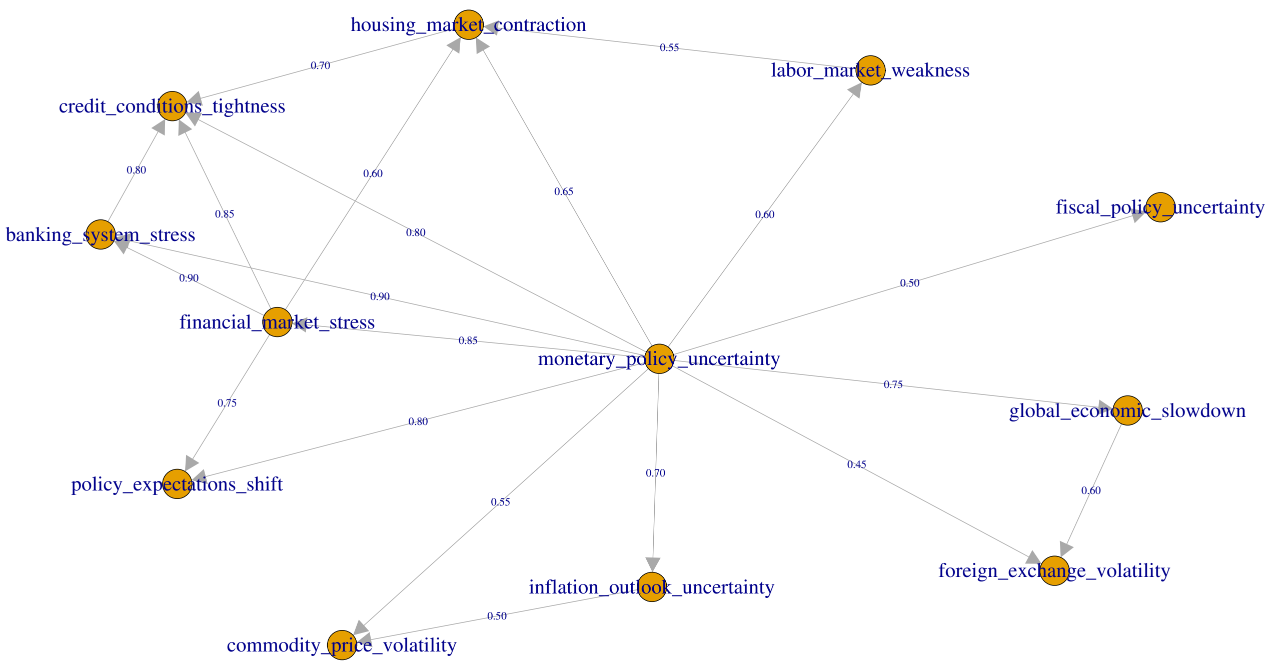

Figure 1: Uncertainty factors from the Federal Reserve minute released on September, 16^{th} 2008 .

Node	Description
monetary policy uncertainty	Uncertainty about the future stance of monetary policy
financial market stress	Stress in short- and long-term financial markets (e.g., repo, bond spreads)
labor market weakness	Weakening employment and payroll trends
housing market contraction	Declining residential investment and home-price pressure
inflation outlook uncertainty	Uncertainty about the path of headline and core inflation
credit conditions tightness	Tight credit conditions for households and businesses
global economic slowdown	Deceleration of growth in advanced and emerging economies
commodity price volatility	Volatility in oil and other commodity prices
banking system stress	Strains on banks and other financial institutions (e.g., Lehman, AIG)
fiscal policy uncertainty	Uncertainty about fiscal actions (tax rebates, GSE conservatorship)
foreign exchange volatility	Movements in the dollar versus major trading partners
policy expectations shift	Changes in market expectations for the future policy path

Table 1
Nodes description (reasoning) from the prompting output of the Federal Reserve minute released on September, 16^{th} 2008

Figure 2 shows the results of a minute released on April 29, 2020, during the early months of the COVID-19 pandemic. The network of uncertainty factors is relatively dense compared to the previous application. Worries related to the then-prevailing pandemic situation were expressed as the second highest edge coming out of *monetary policy uncertainty*, after *financial market uncertainty*. Spillover of the pandemic could be retrieved from the rhetoric used in the minute as being linked to emerging markets, policy communication, and mortgage markets, whose nodes are acting as authorities (nodes receiving many edges). Table 3 explains to elements of Figure 2 and offers a qualitative labeling of nodes regarding uncertainty. Such granular description calls for advanced information retrieval and cannot be performed at the level of topic model and narrative extraction models [7].

In both experiments, the node *monetary policy uncertainty* is a confounder, featuring only outgoing causal edges. Its role, stated as a central node, makes it a hub [37], while some factors were found to be authorities because they attract many nodes with substantial intensities. These factors often interact independently but form further substructures to add more details to the narrative map. In Figure 2, *supply chain uncertainty* is a mediator between *inflation* and *emerging market* uncertainties, denoting inflationist risks related to issues global trade was facing.

The two minutes used were alternatively analyzed using a semantic search model learned on a

Node	Description
Topic 1	'fomc', 'longerrun', 'reopens', 'readings', 'labor', 'pce', 'softness', 'liftoff', 'bounceback', 'fed'
Topic 2	'cares', 'pmccf', 'cpff', 'coronavirus', 'pppIf', 'smccf', 'talf', 'mlf', 'paycheck', 'pdcf'
Topic 3	'pickup', 'damped', 'softness', 'nonfarm', 'homebuilding', 'readings', 'payrolls', 'nondefense', 'downshift', 'crosscurrents'
Topic 4	'kiley', 'reifschneider', 'elb', 'fomc', 'longerrun', 'mertens', 'liftoff', 'laubach', 'labor', 'roberts'
Topic 5	'fomc', 'minneapolis', 'minnesota', 've', 'montana', 'kocherlakota', 'fed', 'fettig', 'll', 'dissents'

Table 2
Top 5 topics uncovered via Top2vec model [10] on the Federal Reserve minute released on September, 16^{th} 2008

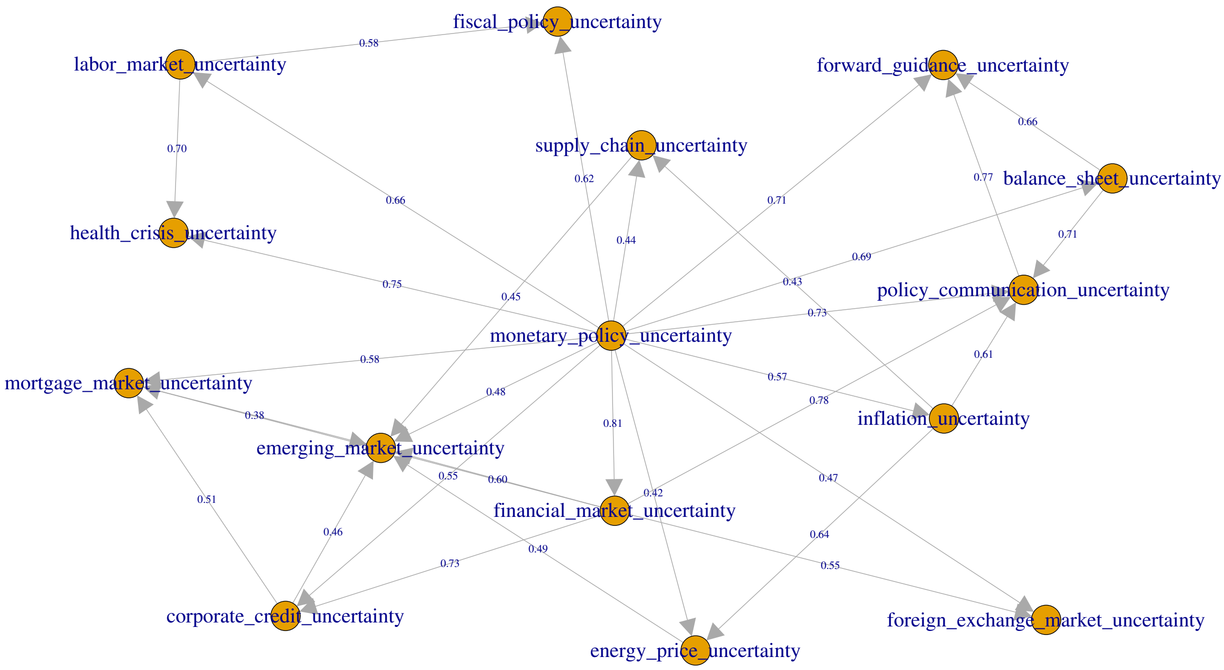

Figure 2: Uncertainty factors from the Federal Reserve minute released on April, 29^{th} 2020 .

Node	Description
monetary policy uncertainty	Stance, tools and timing of Fed policy actions (rate path, asset purchases, balance-sheet size)
health crisis uncertainty	Surrounding the evolution of the COVID-19 pandemic, its severity, and the timing of public-health interventions
fiscal policy uncertainty	Size, composition and duration of fiscal stimulus programmes (e.g., CARES Act, PPP, unemployment benefits)
financial market uncertainty	Market-wide conditions: volatility, liquidity, and the functioning of key funding markets
corporate credit uncertainty	Default risk, rating-watch activity and the availability of financing for non-financial firms
emerging market uncertainty	Capital-outflows, commodity-price exposure and sovereign-debt stress in EM economies
mortgage market uncertainty	Delinquency rates, forbearance extensions and the health of the residential-mortgage market
labor market uncertainty	Job loss dynamics, unemployment duration and the speed of labour-market recovery
inflation uncertainty	Near-term price dynamics given disinflationary demand pressure and deflationary commodity shocks.
supply chain uncertainty	Disruptions to production and logistics caused by lockdowns and travel restrictions.
energy price uncertainty	Oil-price trajectories and their spill-over effects on inflation and corporate earnings.
policy communication uncertainty	Clarity, consistency and forward-guidance of monetary-policy communication
balance sheet uncertainty	Ultimate size, composition and timing of unwind of the Fed's balance-sheet holdings
forward guidance uncertainty	Conditionality and timing embedded in the Fed's forward-guidance framework.
foreign exchange market uncertainty	Dollar funding strains abroad and the effectiveness of swap-line arrangements.

Table 3
Nodes description (reasoning) from the prompting output of the Federal Reserve minute released on April, 29^{th} 2020

corpus of central bank speeches released by the Bank of International Settlements (BIS) [10], as an out-of-sample document projected to the learned embedding space [11]. Tables 2 and 4 show the top 5 topics close to each minute and clearly demonstrate issues pertaining to understanding documents falling into the same category of trained corpora. This means the embedding of the two minutes was not fully adopted by the model, and the embeddings' cosine similarities with the term *uncertainty* are too weak to derive meaningful information.

Node	Description
Topic 1	'pickup', 'damped', 'softness', 'nonfarm', 'homebuilding', 'readings', 'payrolls', 'nondefense', 'downshift', 'crosscurrents'
Topic 2	'adb', 'asian', 'iffo', 'aiib', 'asia', 'cgif', 'mdbs', 'bankable', 'abmi', 'paif'
Topic 3	'ach', 'checks', 'presentment', 'check', 'truncate', 'electronically', 'electronic', 'truncation', 'cheques', 'processing'
Topic 4	'policyrelevant', 'pandemic', 'vaccination', 'pepp', 'vaccinations', 'mutations', 'recalibrated', 'infections', 'vaccine', 'coronavirus'
Topic 5	'jersey', 'puerto', 'rico', 'upstate', 'fairfield', 'syracuse', 'albany', 'queens', 'buffalo', 'brooklyn'

Table 4
Top 5 topics uncovered via Top2vec model [10] on the Federal Reserve minute released on April, 29^{th} 2020

5. Discussion

The use of LLMs to extract uncertainty factors proved to be highly informative and cheaper than automated content analysis. It didn't require labels or hyperparameter setting, as for traditional topic models and other NLP applications. While a rough prompt yielded unstable results, making explicit guidance about the graph structure proved helpful in stabilizing results.

Narrative economics has been known to feature many hard-level tasks in NLP, partly due to the granular tests researchers aim to conduct. These tasks often require complex output, along with human expertise, that considers the economic theory and context information for convincing results. Shortcomings of traditional automated text extraction regularly yield poor results because of missing context information and agnostic training strategies borrowed from statistical learning [38].

LLMs, as deep neural networks, were trained on large amounts of documents, and their ability to explain their results makes them the go-to solution for text analysis. This human-like *reasoning* feature could be considered as a powerful tool at the hand of practitioners whose aim is to maximize information extraction and test context-based results.

In the case of uncertainty, our experiment proved to be effective for a semantic retrieval of latent factors in accordance with monetary policy practice. The DAG-structure of uncertainty determinants was found to be robust in event-driven documents related to the financial crisis or the pandemic. Besides elements that constitute nodes of the graph, edges feature information about the intensity of factors accounting for risks and unclear outlooks, as well as directed edges to track causality within the map. Edges could be used to separate factors from an outlook perspective (forward-looking or backward-looking) or from a positional perspective (internal and/or external factors pertaining to central banks or national/international developments).

Setting a unique central node in the output graph could be seen is a pre-processing step or a normalization process guaranteeing a fair comparability between generated graphs. It helps comparing uncertainty formation and assess the degree of resilience from the edges intensities for further macroprudential simulations.

Further applications of network analysis offer additional storytelling elements, in terms of hubs and authorities, and confirm the importance of having a structural representation able to cope with further tests. Particularly, latent factors could be clustered into cliques [39] and communities [40] to shed light on hidden patterns and their interactions. This brings further insights into LLMs robustness in terms of provided answers and the quality, or richness, of data used to train these models.

In central bank communication, such graph-structure is indicated for comparative analysis [41] and document scaling tasks [42] both at the institutional and member levels. It circumvents agnostic, static models and term-based scaling techniques to offer a theory-relevant basis for document comparison. Running at the document level is one indisputable advantage of this methodology, as it does not depend on the whole corpus and is not influenced by other documents.

Prompts could be augmented with human expertise (labels or examples for few-shot prompting or Chain-of-Thought) to better guide the experiment. This would fit extensive research plans but may come at the risk of misguiding the inference if examples do not cover the depth of the task.

6. Conclusion

Narrative economics established itself as a leading research field in social sciences and witnessed an unprecedented development, correlatively with new NLP tools available to practitioners. The wide adoption of LLMs as a powerful technology to solve difficult tasks helped researchers speed up task execution and enhance data extraction at unprecedented levels. Despite many shortcomings, LLMs have been praised as a reliable, cheaper alternative to traditional automated content analysis, whose implementation features human intervention for difficult tasks.

This work proved the usefulness of LLMs in capturing semantic, event-based uncertainty in selected minutes belonging to monetary policy practice. The output, given as a DAG of uncertainty factors, shapes a storytelling pertaining to different events and contexts witnessed at the Federal Reserve during the past twenty years. Such results did not require extensive prompting strategies but simply setting a central node in the graph output, as a normalization step for an accurate comparison.

The network structure overcomes weaknesses of topic models and other transfer-learning schemes, which are not always pertinent to the field of central bank communication or require large corpora and expensive training to achieve robust results. Latent factors were found to match the prevailing context of the minutes, and their interactions offer additional information to practitioners in terms of causality testing and uncertainty formation. The network linkage may be used to assess uncertainty resilience and serves as a basis for macroprudential simulations, in parallel to its suitability for central bank communication.

Declaration on Generative AI

The authors employed *DeepL* to rectify grammatical errors and enhance the readability of the paper. At the final stage, the authors conducted a thorough review and editing process, taking full responsibility for the publication's content.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, p. 6000–6010. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[2] C. H. Martin, T. Peng, M. W. Mahoney, Predicting trends in the quality of state-of-the-art neural networks without access to training or testing data, Nature Communications 12 (2021). doi:10.1038/s41467-021-24025-8.

[3] E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big?, in: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 2021, pp. 610–623.

[4] J. T. Ornstein, E. N. Blasingame, J. S. Truscott, How to train your stochastic parrot: large language models for political texts, Political Science Research and Methods 13 (2025) 264–281. doi:10.1017/psrm.2024.64.

[5] R. J. Shiller, Narrative Economics, Working Paper 23075, National Bureau of Economic Research, 2017. doi:10.3386/w23075.

[6] B. Kwon, T. Park, F. Perez-Cruz, P. Rungcharoenkitkul, Large language models: a primer for economists, BIS Quarterly Review (2024). URL: https://ideas.repec.org/a/bis/bisqtr/2412b.html.

[7] E. Ash, G. Gauthier, P. Widmer, Relatio: Text semantics capture political and economic narratives, Political Analysis 32 (2024) 115–132.

[8] A. Greenspan, Risk and uncertainty in monetary policy, American Economic Review 94 (2004) 33–40. doi:10.1257/0002828041301551.

[9] S. R. Baker, N. Bloom, S. J. Davis, Measuring economic policy uncertainty*, The Quarterly Journal of Economics 131 (2016) 1593–1636.
[10] S. Diaf, F. Schütze, Uncovering uncertainty in narrative economics: A semantic search approach, in: New Frontiers in Textual Data Analysis, Springer Nature Switzerland, 2024. doi:10.1007/978-3-031-55917-4_26.
[11] D. Angelov, Top2vec: Distributed representations of topics, arXiv preprint arXiv:2008.09470 (2020). URL: https://arxiv.org/pdf/2008.09470.
[12] F. Schütze, Essays on Measuring Uncertainty, Dissertation, Universität Hamburg, Hamburg, 2025. URL: https://ediss.sub.uni-hamburg.de/handle/ediss/11490.
[13] S. Diaf, F. Schütze, Multilingual monetary policy: Unfolding language and policy preferences of swiss central bankers, in: CARMA Conference 2024, Editorial Universitat Politècnica de València, 2024. doi:10.4995/CARMA2024.2024.17405.
[14] S. Eshima, K. Imai, T. Sasaki, Keyword assisted topic models, ArXiv abs/2004.05964 (2020). URL: https://arxiv.org/pdf/2004.05964.
[15] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, 2013. URL: https://arxiv.org/pdf/1301.3781.
[16] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 2019, pp. 4171–4186. doi:10.18653/v1/N19-1423.
[17] F. Xu, Q. Hao, C. Shao, Z. Zong, Y. Li, J. Wang, Y. Zhang, J. Wang, X. Lan, J. Gong, T. Ouyang, F. Meng, Y. Yan, Q. Yang, Y. Song, S. Ren, X. Hu, J. Feng, C. Gao, Y. Li, Toward large reasoning models: A survey of reinforced reasoning with large language models, Patterns 6 (2025) 101370. doi:10.1016/j.patter.2025.101370.
[18] N. Baerg, C. Binder, Automated detection of emotion in central bank communication: A warning, National Institute Economic Review 269 (2024) 82–91. doi:10.1017/nie.2024.31.
[19] A. Zarifhonarvar, Generating inflation expectations with large language models, Journal of Monetary Economics 157 (2026) 103859. doi:10.1016/j.jmoneco.2025.103859.
[20] J. M. Echterhoff, Y. Liu, A. Alessa, J. McAuley, Z. He, Cognitive bias in decision-making with LLMs, in: Y. Al-Onaizan, M. Bansal, Y.-N. Chen (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2024, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 12640–12653. doi:10.18653/v1/2024.findings-emnlp.739.
[21] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al., Chain-of-thought prompting elicits reasoning in large language models, Advances in neural information processing systems 35 (2022) 24824–24837.
[22] M. D. Packard, P. L. Bylund, B. B. Clark, Keynes and Knight on uncertainty: peas in a pod or chalk and cheese?, Cambridge Journal of Economics 45 (2021) 1099–1125.
[23] J. M. Keynes, The general theory of employment, interest and money, Macmillan, London, 1936.
[24] F. H. Knight, Risk, uncertainty and profit, Houghton Mifflin, Boston, 1921.
[25] B. S. Bernanke, Monetary policy under uncertainty, Speech 327, Board of Governors of the Federal Reserve System (U.S.), 2007. URL: https://ideas.repec.org/p/fip/fedgsq/327.html.
[26] T. Loughran, B. McDonald, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, The Journal of Finance 66 (2011) 35–65.
[27] L. F. Husted, J. H. Rogers, B. Sun, Monetary Policy Uncertainty, International Finance Discussion Papers 1215, Board of Governors of the Federal Reserve System (U.S.), 2017. doi:10.17016/IFDP.2017.1215.
[28] L. Husted, J. Rogers, B. Sun, Monetary policy uncertainty, Journal of Monetary Economics 115 (2020) 20–36.
[29] A. Ito, M. Sato, R. Ota, A novel content-based approach to measuring monetary policy uncertainty using fine-tuned llms, Finance Research Letters 75 (2025) 106832. doi:10.1016/j.frl.2025.106832.
[30] E. C. Arbatli Saxegaard, S. J. Davis, A. Ito, N. Miake, Policy uncertainty in japan, Journal of the

Japanese and International Economies 64 (2022) 101192. doi:10.1016/j.jjie.2022.101192.
[31] P. Parschan, C. Jakob, Computational measurement of political positions: A review of text-based ideal point estimation algorithms, Quality & Quantity (2025). URL: https://link.springer.com/article/10.1007/s11135-025-02500-4.
[32] A. Korinek, Generative ai for economic research: Use cases and implications for economists, Journal of Economic Literature 61 (2023) 1281–1317. doi:10.1257/jel.20231736.
[33] M. Pfeifer, V. P. Marohl, Centralbankroberta: A fine-tuned large language model for central bank communications, The Journal of Finance and Data Science 9 (2023) 100114. doi:10.1016/j.jfds.2023.100114.
[34] Y. Gorodnichenko, T. Pham, O. Talavera, The voice of monetary policy, American Economic Review 113 (2023) 548–84. doi:10.1257/aer.20220129.
[35] M. E. Bontempi, W. Charemza, S. Makarova, Economic uncertainty measures, experts and large language models, Journal of International Money and Finance 157 (2025) 103369. doi:10.1016/j.jimonfin.2025.103369.
[36] OpenAI, gpt-oss-120b & gpt-oss-20b model card, 2025. URL: https://arxiv.org/abs/2508.10925. arXiv:2508.10925.
[37] J. M. Kleinberg, Hubs, authorities, and communities, ACM computing surveys (CSUR) 31 (1999) 5–es.
[38] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[39] I. Derényi, G. Palla, T. Vicsek, Clique percolation in random networks, Phys. Rev. Lett. 94 (2005) 160202.
[40] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment (2008) P10008.
[41] S. Diaf, Policy preference at central banks: Quantifying monetary policy signals using keyword topic models, WiSo-HH working paper series 69 (2022). URL: https://d-nb.info/1273646363/34.
[42] S. Diaf, U. Fritsche, TopicShoal: Scaling partisanship using semantic search, in: Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022), KONVENS 2022 Organizers, Potsdam, Germany, 2022, pp. 167–174. URL: https://aclanthology.org/2022.konvens-1.20.