Issues in Information Systems
Volume 26, Issue 2, pp. 96-108, 2025

DOI: https://doi.org/10.48009/2_iis_108

Assessing the Alignment of FOMC Statements with Minutes using Large Language Models

Adam Patterson, *University of Connecticut, adam.patterson@uconn.edu*

Abstract

This paper pioneers a quantitative and AI-driven framework for evaluating central bank communication. The Federal Open Market Committee (FOMC) releases an official statement when meetings conclude and monetary policy is being announced, and minutes of the meeting three weeks after. The statement acts as a summary of the meeting whereas minutes provide an overview of meeting discussions. The former is contained within one to two pages while the latter covers approximately ten pages. This study leverages Amazon Nova, a large language model (LLM) on Amazon Bedrock in AWS, to evaluate the accuracy and completeness of the statements relative to the minutes. Abstractive summaries are created from the full minutes and semantically compared to the official FOMC statements using SBERT-based embeddings. By computing and analyzing these similarity scores, this study assesses if FOMC statements sufficiently capture the key themes and policy discussions reflected in the minutes. Given the significant financial market movements immediately following FOMC statements due to the rise of automated trading, ensuring transparency in central bank communication is critical for market efficiency. Findings indicate that FOMC documents consistently display general alignment, showcasing stability over time, with no discernible heterogeneity by year or season. Only two observations (n=57) qualify as statistical outliers. These findings are robust to multiple iterations of abstractive summarization.

Keywords: monetary policy, central bank, large language models, transparency, consistency

Introduction

The FOMC, or committee, plays a pivotal role in shaping U.S. monetary policy, directly impacting financial markets, economic growth, and overall societal well-being. Tasked with advancing Congress's dual mandate of price stability and maximum employment, the committee's decisions have far-reaching implications (Thornton, 2012). Comprised of twelve members, the committee meets eight times per annum to assess monetary policy given domestic and international economic conditions. Each meeting concludes with a decision on the federal funds target rate and the release of a statement summarizing key discussions. Approximately three weeks later, the committee publishes meeting minutes, providing a more detailed account of the deliberations. A full transcript of the meeting is released five years after the statement.

Investors, policymakers, and analysts rely on FOMC statements to understand the target range for the federal funds rate, interpret the committee’s perspectives, and anticipate future policy decisions. The alignment of the statements with discussions from the meeting is essential for efficient markets, as the statement is the first text communicated publicly. This is increasingly relevant given the massive rise of AI-powered trading algorithms, which rely on textual data, such as FOMC statements. AI-driven technologies have been progressively leveraged to aid the decision-making of various investment funds (Parisi & Manaog, 2025; Xie et al., 2023). Moreover, the development and application of algorithmic trading will continue to grow as research underscores the enhanced accuracy of LLMs in financial market prediction (Kirtac & Germano, 2024).

As algorithmic trading systems increasingly depend on FOMC statements, inaccuracies in these texts could influence trading decisions, exacerbate market volatility, and inefficiently allocate capital (Guida, 2018). If these statements misrepresent key economic discussions from meetings, machine-driven financial decisions could be based on incomplete or misleading information. In addition to AI trading systems, if statements omit key information or misrepresent committee conversations, they could distort market expectations and policy interpretations for all stakeholders. Given these high stakes, it is surprising that rigorous evaluations of the accuracy of FOMC statements remain rare.

Despite their significance, only one systematic study, with many limitations, has quantitatively assessed the alignment of FOMC statements to the full meeting minutes (Patterson, 2024). Prior research has explored the financial market impact of FOMC communications but has largely assumed that the statements accurately reflect the underlying discussions. This study introduces a novel, AI-powered framework to evaluate the completeness of FOMC statements by leveraging an LLM for abstractive summarization of minutes and then calculating BERT-based sentence embeddings to capture semantic similarity between the abstractive summary and the statement.

By applying these techniques, the author aims to examine how closely the FOMC statements align with the full minutes. Additionally, exploring heterogeneity by season and year is of interest to the researcher. If discrepancies exist, the findings in this study could spur policy reforms in central bank communications, influence how journalists and analysts interpret FOMC releases, or provide a basis for investors to adjust how they weigh FOMC summaries in forecasting models.

The alignment of FOMC statements to minutes is more critical than ever. This study provides a data-driven foundation for the development of AI trading algorithms and for policymakers, financial analysts, and researchers to evaluate the effectiveness of monetary policy communication. This framework helps ensure that central bank messaging remains clear, transparent, and aligned with reality. Unlike transcript-based methods that require a five-year wait, this methodology enables a more immediate assessment of transparency, just three weeks after policy announcement.

The findings in this study provide institutions with quantitative measures to promote information symmetry and improve how policy announcements are priced into markets. Beyond its application to FOMC communication, this research establishes a scalable framework that can also be extended to other government communications, such as Congressional hearings, SEC filings, and IMF reports, for investigating potential biases in official disclosures.

Literature Review

The committee policy announcement, contained in the statement, consistently moves financial markets, as evidenced by significant increases in volatility and trading volume upon release of the statement (Farka & Fleissig, 2013; Rosa, 2016). These effects are particularly pronounced among firms specializing in automated trading, which react rapidly to perceived mispricings from the text (Harkrader & Weitz, 2020). Additionally, prior statements influence pre-meeting positioning, as Lucca and Moench (2015) show that U.S. equities experience statistically significant excess returns ahead of scheduled meetings. This suggests that market actors form expectations based on earlier communications (Manfre, 2024).

Numerous studies have analyzed FOMC statements using natural language processing (NLP) tools to interpret or predict market reactions (Huang & Kuan, 2021; Osowska & Wójcik, 2024; Rosa, 2013). However, much of this research implicitly assumes that these statements accurately capture the full scope of discussions held during the committee meetings. Emerging evidence challenges this assumption.

Fischer et al. (2023) demonstrate that immediate public access to full meeting transcripts would improve policy expectation forecasts, implying that statements omit valuable information. Similarly, Apel (2022) finds that transcripts, unlike minutes, contain forward-looking content relevant to macroeconomic variables. Tadle (2022) and Venade & Grilo (2024) show that minutes contain additional sentiment or signals that influence asset prices beyond what the statements convey. These findings collectively suggest that statements may not fully reflect the substance of FOMC discussions.

This discrepancy may stem in part from the stylized, cautious nature of official statements. Prior work has shown that FOMC communications follow a templated format, avoid emotional language, and are subject to conformity pressures (Kim et al., 2024; Mazis & Tsekrekos, 2017). Hansen (2018) finds that public release requirements, such as minutes and transcripts, can make policymakers more cautious, consistent with the Hawthorne effect (Adair, 1984). Acosta (2023) confirms increasing conformity in language following transparency mandates. Taskin and Akal (2025) further find that even during unprecedented events like the COVID-19 pandemic, the tone of FOMC statements remained unusually stable. These patterns highlight the potential for meaningful divergence between actual discussions and their public summaries.

By leveraging Natural Language Processing (NLP) techniques and LLMs, this research quantifies the extent to which FOMC statements capture the key themes, risks, and policy discussions present in the minutes. Traditional NLP algorithms often ignore context, whereas deep learning NLP algorithms, such as LLMs, take into account the semantic meaning of the text (Pfeifer & Marohl, 2023). Thus, LLMs offer human-quality text summarization with machine scalability (Shakil et al., 2024; Zhang et al., 2024). According to Van Veen et al. (2024), LLMs can even outperform field experts in text summarization. In addition, Pfeifer (2023) notes the absence of large language models in the central bank communications literature, further highlighting the novelty of this approach.

To date, only one study (Patterson, 2024) has attempted to systematically quantify the alignment between FOMC statements and minutes using transformer-based models. However, this approach is constrained by a 512-token limit on input length and relies on token-level embeddings from BERT to assess cosine similarity, which primarily captures lexical overlap rather than deeper semantic alignment. Moreover, it does not address the inherent variability in outputs from LLMs. This study builds on Patterson's foundation but overcomes its methodological limitations in several ways. First, it employs an LLM to generate abstractive summaries of FOMC minutes that are unconstrained by token limits, as all minutes are less than 128 thousand tokens. Second, it uses Siamese BERT-Networks (SBERT) to calculate sentence embeddings and cosine similarity, allowing for a more nuanced comparison of semantic meaning between documents. Third, it incorporates a robustness check to account for output variability in LLM summarization, which was overlooked in prior work. By addressing these gaps, this research provides a more reliable and scalable framework for assessing the fidelity of central bank communications.

Methodology

The scope of this research includes the entirety of meetings with the Honorable Jerome Powell serving as Chairman, serving as a control for bespoke communication preferences specific to the chair. Committee documents were web scraped directly from the Federal Reserve website using *beautiful soup*, *url*, and *requests* packages in Python. Minutes and statements from March 2018 through March 2025 were collected in HTML format for analysis.

Analysis used in this investigation utilizes Amazon Nova Micro in Amazon Bedrock on Amazon Web Services (AWS) to generate abstractive summaries of the meeting minutes. Amazon Nova was chosen for its exceptional pricing efficiency compared to other foundational models (Amazon Web Services, 2025a). Moreover, Amazon Nova Micro exhibits state-of-the-art language understanding and reasoning, which are crucial for accurately capturing the nuances of economic and policy discussions (Amazon Web Services, 2025b). A key advantage of Nova is its 128k maximum token length, which allows for the entire content of the FOMC minutes to be used as input for the LLM, thereby overcoming the 512-token limitation of previous research established by Patterson (2024). This ensures that the generated summaries are comprehensive and reflect all key information from the original minutes.

Nova was prompted systematically for each meeting minutes using the following instruction:

> *You are an expert editor summarizing economic information.*
> *Rewrite the following text into a concise, 700 token abstractive summary.*
> *Focus on clarity, brevity, and essential insights while avoiding unnecessary details.*
> *Do not include any information about dates (month/year).*

A significant benefit of using Nova is that no data preprocessing steps, such as tokenization or embedding creation, were required from the researcher. The LLM internally handles these processes, tokenizing the text, creating embeddings, predicting embeddings based on cosine similarity, converting the predicted embeddings to tokens, and then converting the tokens to output text in the generated abstractive summaries. Overall, the internal data processing streamlines the analysis and ensures consistency in data handling.

LLM outputs are inherently stochastic as the model samples from a probability distribution over the vocabulary during token generation. As a result, the same input can yield different summaries depending on decoding strategies, such as top-k sampling and adjusting the temperature parameter. Temperature controls the randomness of LLM output during text generation by scaling the probability distribution before sampling. A model with a temperature of zero always selects the most likely next token. While setting the temperature to zero reduces randomness by forcing the model to select the highest-probability token at each step, full reproducibility is not guaranteed in practice due to potential nondeterminism in backend infrastructure, hardware-level operations, or versioning. At the time of this research, Nova did not support a temperature parameter.

Following summary generation, SBERT is employed to evaluate semantic text similarity between pairwise statements and abstractive summaries (Sun et al., 2022). SBERT is a modified model based on BERT that uses transformers to derive semantically meaningful sentence embeddings for comparing text similarity (Reimers & Gurevych, 2019). SBERT tokenizes and calculates embeddings for each sentence. SBERT truncates sentences to the 512-token limit of BERT in creating embeddings, which is not a concern for evaluating text in this study. Multiple sentences are encoded at once using a batch model named sentence-transformers, generating semantically meaningful sentence embeddings for the entirety of texts (Lamsal et al., 2024).

Cosine similarity is then utilized to measure the proximity of pairwise sentence embeddings and thus understand the semantic similarity between texts. Cosine similarity provides a quantitative metric of semantic alignment between the abstractive summary and official minutes by measuring the closeness of the two vectors. Scores are bounded by negative one and one. A score of one denotes identical vectors (perfect semantic alignment), a negative one indicates two opposite vectors (contradictory semantic meaning), whereas zero specifies an orthogonal relationship between texts (no semantic relationship). These scores form the basis for evaluating the completeness and accuracy of FOMC statements in reflecting the underlying discussions.

The full python code used for this analysis is posted on the author's Github account (here) to promote transparency and act as a foundation for future research. The code contains webscrapping, calling the Bedrock client, prompting the LLM to summarize documents, utilizing SBERT to gain semantic similarity scores, creating visuals, and running the robustness check.

Results

The key findings from the novel application of LLMs are presented, evaluating the alignment in FOMC communications and providing insights on the consistency and transparency of central bank messaging. When a statement and its corresponding meeting minutes exhibit a cosine similarity of one, it indicates that the committee conveyed the same information and sentiment, using very similar language. This is the ideal outcome for maximum credibility. A cosine similarity of zero implies that the documents are completely unrelated in their content. This would be a highly unusual and problematic finding, indicating a severe disconnect in communication. A cosine similarity of negative one suggests that the FOMC statement and meeting minutes convey contrary or opposing messages. For example, one might be highly optimistic while the other is highly pessimistic about the same economic indicators. This would be an extreme case of misalignment and a serious concern for central bank communication.

Although there are no universal thresholds for cosine similarity that apply across all domains, the following interpretations are used as criteria in this research. Similarity scores greater than 0.75 indicate that the two documents are strongly aligned, suggesting that the statement mirrors the sentiment within the minutes. A similarity score within the range of 0.50 to 0.75 signifies that the texts are generally aligned. Scores within this range exhibit significant shared semantic alignment. A similarity score between 0.20 and 0.50 suggests weak alignment, exhibiting few thematic commonalities. Values between 0.0 and 0.20 represent very low to no alignment. Negative similarity scores indicate misalignment, with larger negative values supporting stronger misalignment. This would suggest that the texts convey contradictory themes and semantic meaning.

Table 1 presents the descriptive statistics of similarity scores for all pairwise meetings, while Figure 1 illustrates their distribution. A mean similarity score of 0.579 falls into the generally aligned threshold established for this research. This result suggests that the core message communicated in FOMC statements are generally reflected and supported by the abstractive summarization of the more detailed discussion minutes. The standard deviation of 0.036 indicates a precise result, as a majority of meetings fall within the aligned category. Figure 1 visually illustrates the distribution of these similarity scores, providing a clear representation of their concentration. The dashed red line indicates the mean whereas the dashed green line represents the median.

Table 1. Similarity Score Descriptive Statistics

N	Mean	Median	Standard Deviation
57	0.579	0.585	0.036

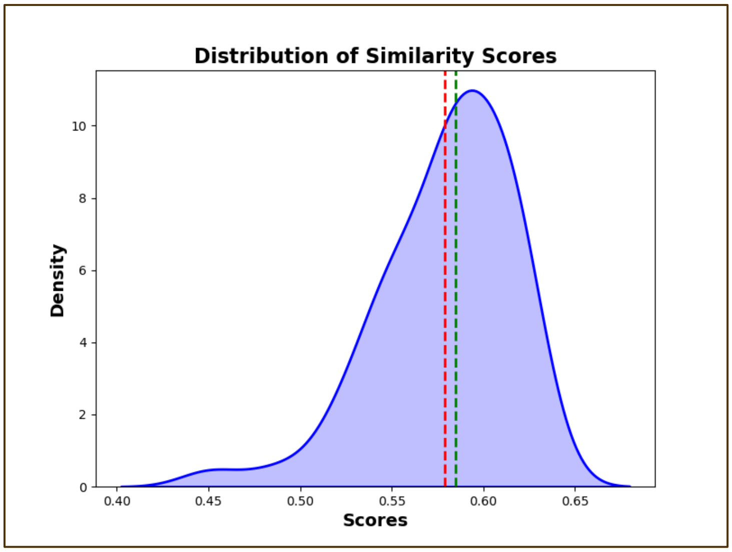

Figure 1. Distribution of Similarity Scores

Findings indicate two statistical outliers using the IQR method, suggesting possible room for increased transparency in FOMC statements. Evaluating potential drivers of the outlier events is beyond the scope of this investigation. Figure 2 displays a scatterplot mapping each meeting to its corresponding similarity score, providing a visual analysis of the depth of outliers. Additionally, it is noted that (a) scores appear to get tighter with time and (b) a wave-like trend establishes itself. The researcher ponders if this wave-like pattern exhibits seasonal variation. To evaluate heterogeneity in the similarity score, results are grouped by year or season. Figure 3 showcases the average similarity score by year, whereas Figure 4 displays the average similarity score by season. Insights generated reveal a slight deviation in variation by year, with no noticeable trend and a relatively homogenous result across seasons, as the four median values are approximately equal. These figures indicate a lack of significant year and season heterogeneity in FOMC communication strategies.

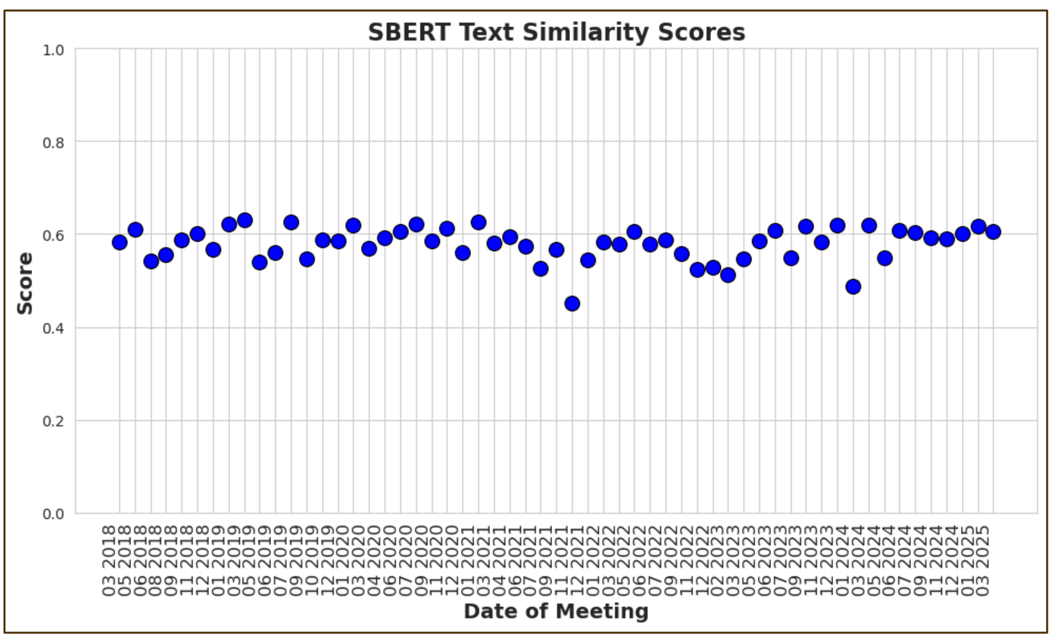

Figure 2. Scatterplot of Scores and Meeting Dates

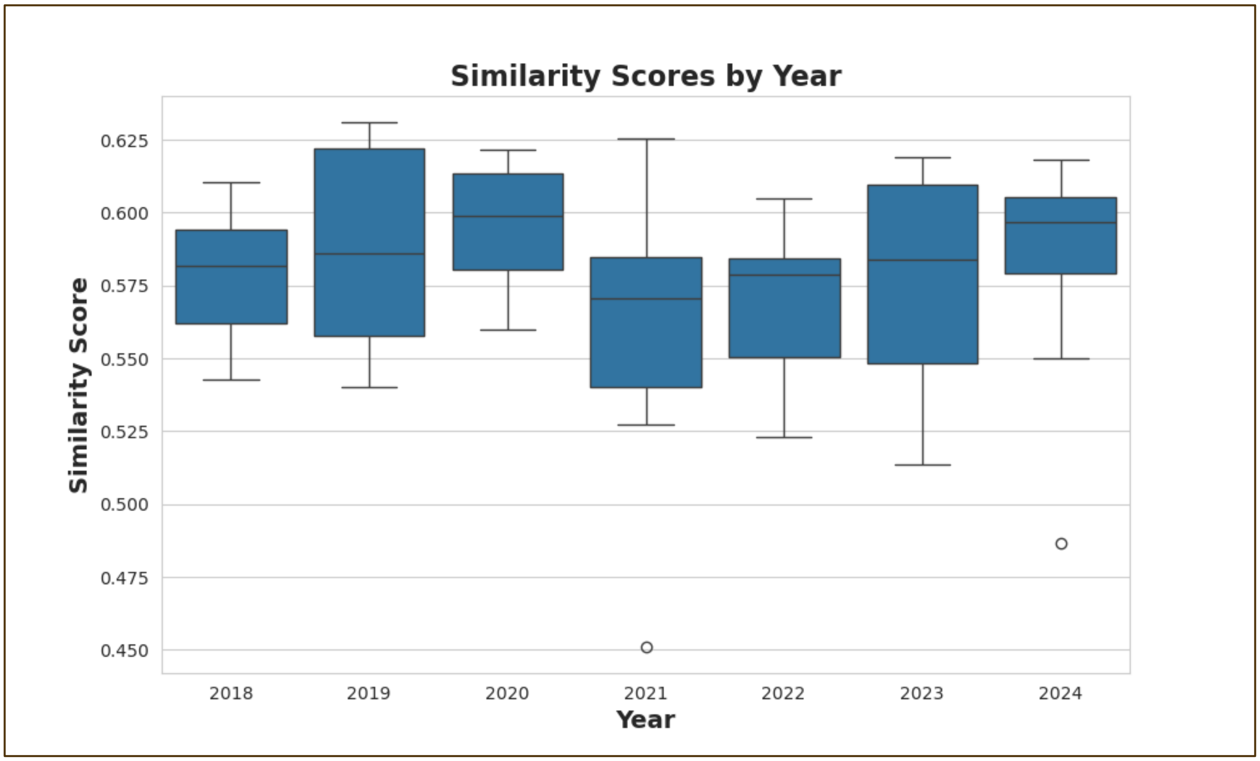

Figure 3. Similarity Score by Year

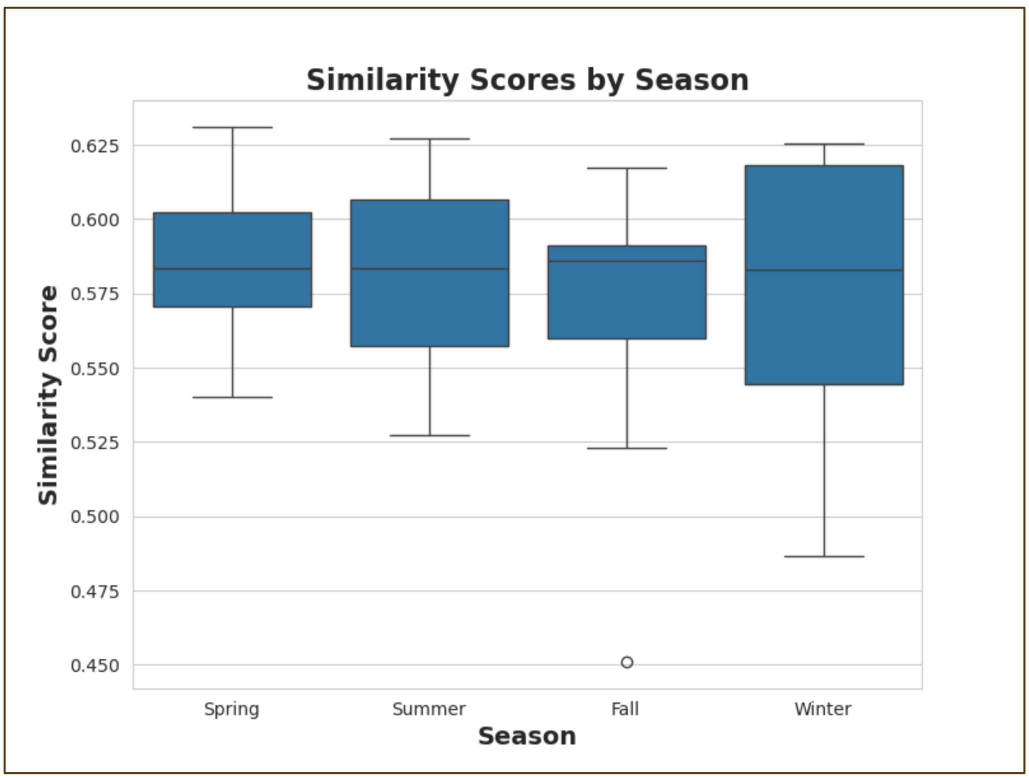

Figure 4. Similarity Score by Season

Robustness Analysis
Given the risk of generating misleading or fabricated information, known as hallucinations, and the stochastic nature of LLMs' output, it is essential to verify the consistency of similarity score evaluations as a robustness check (Du et al., 2024). A random sample of two abstractive summaries was manually examined by the author and provided no evidence of hallucination. Furthermore, unlike other foundational models, Amazon Nova Micro in Amazon Bedrock does not currently allow for direct control over temperature settings, introducing potential increased variability in generated summaries.

To ensure the stability of findings, a robustness check is conducted by running 25 iterations of the full evaluation process. 25 was chosen instead of 100 to reduce cost and throttling issues. By comparing the distribution of similarity scores across iterations, we assess whether our conclusions remain valid despite inherent randomness in LLM-generated summaries. This analysis strengthens the reliability of the results and confirms that any observed patterns in FOMC summary alignment are not artifacts of a single model output but rather robust across multiple runs.

Table 2 displays the descriptive statistics of similarity scores for 25 iterations, while Figure 5 illustrates their distribution. The dashed red line exhibits the mean. The average number of outliers is 1.08 per iteration. The mean of the mean similarity scores from the 25 iterations is very close to the original finding. The difference of .006 is an encouraging result to validate the findings of this research. In addition, the standard deviation of 0.040 indicates stability within the iterations. This consistency suggests that the observed similarity between statements and minutes is not the result of random variation or stochastic LLM output but rather reflects a systematic relationship of alignment between texts.

Table 2. Bootstrapped Similarity Score Descriptive Statistics

N	Mean	Median	Standard Deviation
25	0.573	0.578	0.040

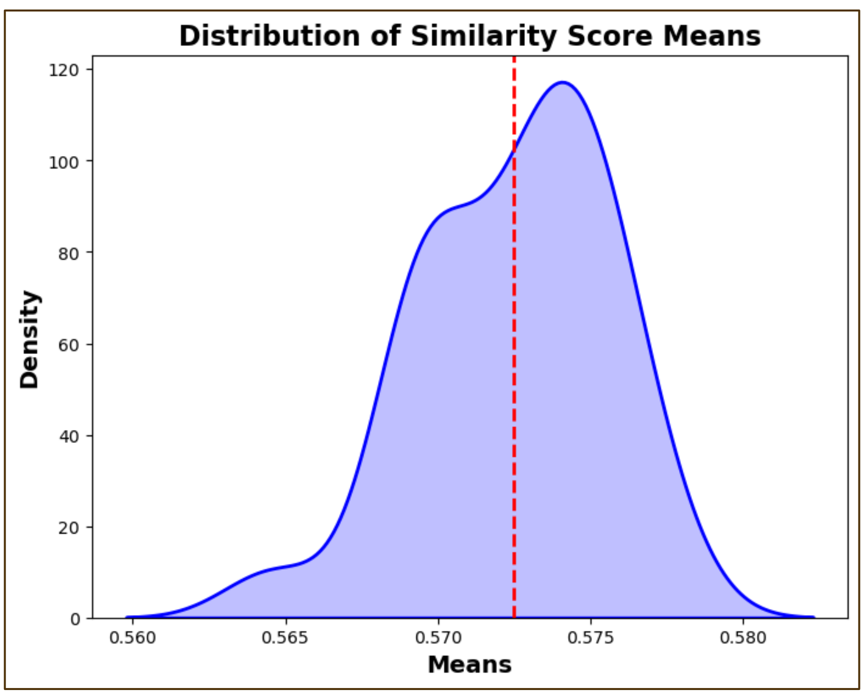

Figure 5. Distribution of Means

Discussion

This study introduces an innovative AI-driven approach to quantitatively assess the alignment between FOMC statements and the subsequent meeting minutes. Utilizing the Amazon Nova LLM to generate abstractive summaries of the minutes and comparing them semantically to the official statements with SBERT embeddings, this study provides a data-backed framework for evaluating central bank communication. The analysis of similarity scores reveals an average score of 0.579, demonstrating general alignment between documents, with minimal yearly and seasonal heterogeneity. Although the wording between texts may vary, they have significant overlap of thematic content and overall sentiment. They generally state the same conditions and point in the same direction.

Notably, the identification of only two statistical outliers suggests a general consistency in FOMC communication, though it highlights potential instances where greater transparency could be warranted. The robustness check, involving 25 iterations, further validates the stability of these findings despite the inherent variability in LLM-generated summaries. In essence, the analysis suggests that the FOMC is largely successful in maintaining fairly strong consistency between its public announcements and underlying discussions. This methodology offers a valuable tool for ongoing monitoring of the FOMC's public messaging.

Conclusion

This research demonstrates the feasibility and potential of using LLMs to evaluate the alignment between FOMC statements and meeting minutes. The quantitative framework developed reveals a consistent alignment between FOMC statements and minutes, with an average similarity score of 0.579 and a standard deviation of 0.036. The high precision of the initial estimates are corroborated by the robustness check and provides increased validity in the study's findings. Overall, the similarity scores are consistent in aggregate, and between years and seasons. This result provides confidence in the central bank's communication strategies. However, the identification of two outlier events indicates that there may be room for improvement in the consistency of FOMC communications, although given the sample of 57, two is a relatively small number (< 4%). The AI-driven approach utilized in this study provides a scalable framework for assessing central bank communication, with potential applications for other government institutions. Ultimately, this methodology contributes to ensuring transparency, fidelity, and accountability in central bank messaging, which is crucial for market efficiency and stability.

Limitations

As of when this research was performed, Amazon Nova models do not provide the ability to adjust temperature. Setting the temperature to zero would allow less variability in the abstractive summaries and reduce stochasticity. A robustness check was conducted in an attempt to account for this limitation. A more expensive foundation model, with temperature settings, could be utilized to account for this limitation or to compare with the results from Nova. Thus, it is important to note that the provided project code will not reproduce the exact results published in this paper. No LLM will provide perfect reproducibility as they are stochastic models. Additionally, no fine-tuning or knowledge bases were used as part of this analysis. The stock version of Amazon Nova's foundation model was deployed as-is to generate abstractive summarization. It may be of interest to fine-tune a model using stated FOMC objectives and published text.

Future Research and Conclusions

There are many avenues for extending this work into future research. One potential extension is to use SBERT to evaluate text similarity metrics between the FOMC published statement and minutes for each meeting, and to then compare similarity metrics between LLM-generated statements and the minutes. If the LLM-generated statements demonstrate higher alignment with the full FOMC minutes than the official statements, important questions and insights about the role of AI in high-stakes policy communication may be surfaced. It may show that AI-generated reports enhance the accuracy of public-facing policy documents, establishing AI-based auditing as a critical tool for government and institutional transparency.

It would also be interesting to map alignment metrics to variation in the Fed Fund futures, two-year and ten-year treasury yields, S&P 500 index, VIX, and the dollar index. Exploring how financial assets react to similarity metrics at different lagged, leading, or coincidental periods may surface insights into the severity of misallocated capital. Furthermore, future research may deep dive into the dates of outliers and examine potential factors driving the misalignment. Additionally, the robustness analysis may be improved upon to determine if the outliers in each iteration are homogenous in terms of meeting date.

The full code used for this research, found within the GitHub link embedded in the Methodology section, may act as a starting point for researchers to build upon this study. The files may also provide a framework for evaluating communication from other institutions or central banks. Furthermore, the code base may be easily mapped to various financial assets for analyzing movement around FOMC meeting dates conditional on document alignment.

References

Acosta, M. (2023). A new measure of central bank transparency and implications for the effectiveness of monetary policy. *International Journal of Central Banking*, *19*(3), 49–97.

Adair, J. G. (1984). The Hawthorne Effect: A Reconsideration of the Methodological Artifact. *Journal of Applied Psychology*, *69*(2), 334.

Amazon Web Services. (2025a). *Amazon Bedrock Pricing*. https://aws.amazon.com/bedrock/pricing/

Amazon Web Services. (2025b). *Amazon Nova: Generative AI on AWS*. https://aws.amazon.com/ai/generative-ai/nova/

Apel, M., Blix Grimaldi, M., & Hull, I. (2022). How Much Information Do Monetary Policy Committees Disclose? Evidence from the FOMC's Minutes and Transcripts. *Journal of Money, Credit and Banking*, *54*(5), Article 5. https://doi.org/10.1111/jmcb.12885

Du, X., Xiao, C., & Li, Y. (2024). HaloScope: Harnessing Unlabeled LLM Generations for Hallucination Detection. *Advances in Neural Information Processing Systems*, *37*, 102948–102972.

Farka, M., & Fleissig, A. R. (2013). The impact of FOMC statements on the volatility of asset prices. *Applied Economics*, *45*(10), Article 10. https://doi.org/10.1080/00036846.2011.615732

Fischer, E., McCaughrin, R., Prazad, S., & Vandergon, M. (2023). *Fed Transparency and Policy Expectation Errors: A Text Analysis Approach* (SSRN Scholarly Paper No. 4643656; Issue 4643656). Social Science Research Network. https://doi.org/10.2139/ssrn.4643656

Guida, T. (2018). *Big Data and Machine Learning in Quantitative Investment*. John Wiley & Sons.

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach*. *The Quarterly Journal of Economics*, *133*(2), Article 2. https://doi.org/10.1093/qje/qjx045

Harkrader, J. C., & Weitz, D. J. (2020). *How Do Principal Trading Firms and Dealers Trade around FOMC Statement Releases?* https://www.federalreserve.gov/econres/notes/feds-notes/how-do-principal-trading-firms-and-dealers-trade-around-fomc-statement-releases-20201231.html

Huang, Y.-L., & Kuan, C.-M. (2021). Economic prediction with the FOMC minutes: An application of text mining. *International Review of Economics & Finance*, *71*, 751–761. https://doi.org/10.1016/j.iref.2020.09.020

Kim, W., Spörer, J. F., & Handschuh, S. (2024). *Analyzing FOMC Minutes: Accuracy and Constraints of Language Models* (No. arXiv:2304.10164; Issue arXiv:2304.10164). arXiv. https://doi.org/10.48550/arXiv.2304.10164

Kirtac, K., & Germano, G. (2024). Sentiment trading with large language models. *Finance Research Letters*, *62*, 105227. https://doi.org/10.1016/j.frl.2024.105227

Lamsal, R., Read, M. R., & Karunasekera, S. (2024). CrisisTransformers: Pre-trained language models and sentence encoders for crisis-related social media texts. *Knowledge-Based Systems*, *296*, 111916. https://doi.org/10.1016/j.knosys.2024.111916

Lucca, D. O., & Moench, E. (2015). The Pre-FOMC Announcement Drift. *The Journal of Finance*, *70*(1), Article 1. https://doi.org/10.1111/jofi.12196

Manfre, C. L. (2024, January 1). *FOMC Announcements and Market Reactions. | EBSCOhost*. https://openurl.ebsco.com/contentitem/gcd:179082727?sid=ebsco:plink:crawler&id=ebsco:gcd:179082727

Mazis, P., & Tsekrekos, A. (2017). Latent semantic analysis of the FOMC statements. *Review of Accounting and Finance*, *16*(2), Article 2. https://doi.org/10.1108/RAF-10-2015-0149

Osowska, E., & Wójcik, P. (2024). Predicting the reaction of financial markets to Federal Open Market Committee post-meeting statements. *Digital Finance*, *6*(1), 145–175. https://doi.org/10.1007/s42521-023-00096-8

Parisi, L., & Manaog, M. L. (2025). Optimal Machine Learning- and Deep Learning-driven algorithms for predicting the future value of investments: A systematic review and meta-analysis. *Engineering Applications of Artificial Intelligence*, *142*, 109924. https://doi.org/10.1016/j.engappai.2024.109924

Patterson, A. (2024). Examining Text Consistency Amongst FOMC Statements and Minutes Subtext using Deep Learning. *Issues In Information Systems*. https://doi.org/10.48009/3_iis_2024_120

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*, 100114. https://doi.org/10.1016/j.jfds.2023.100114

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* (No. arXiv:1908.10084). arXiv. https://doi.org/10.48550/arXiv.1908.10084

Rosa, C. (2013). *The Financial Market Effect of FOMC Minutes* (SSRN Scholarly Paper No. 2378398; Issue 2378398). Social Science Research Network. https://papers.ssrn.com/abstract=2378398

Rosa, C. (2016). Walking on thin ice: Market quality around FOMC announcements. *Economics Letters*, *138*, 5–8. https://doi.org/10.1016/j.econlet.2015.10.029

Salah, A., & Hayette, G. (2025). A meta-analysis of supervised and unsupervised machine learning algorithms and their application to active portfolio management. *Expert Systems with Applications*, *271*, 126611. https://doi.org/10.1016/j.eswa.2025.126611

Shakil, H., Farooq, A., & Kalita, J. (2024). Abstractive text summarization: State of the art, challenges, and improvements. *Neurocomputing*, *603*, 128255. https://doi.org/10.1016/j.neucom.2024.128255

Sun, X., Meng, Y., Ao, X., Wu, F., Zhang, T., Li, J., & Fan, C. (2022). Sentence Similarity Based on Contexts. *Transactions of the Association for Computational Linguistics*, *10*, 573–588. https://doi.org/10.1162/tacl_a_00477

Tadle, R. C. (2022). FOMC minutes sentiments and their impact on financial markets. *Journal of Economics and Business*, *118*, 106021. https://doi.org/10.1016/j.jeconbus.2021.106021

Taskin, B., & Akal, F. (2025). Tales of Turbulence: BERT-based Multimodal Analysis of FED Communication Dynamics Amidst COVID-19 Through FOMC Minutes. *Computational Economics*, *65*(1), 117–146. https://doi.org/10.1007/s10614-023-10533-w

Thornton, D. L. (2012). The Dual Mandate: Has the Fed Changed Its Objective? *Review*, *94*(2). https://doi.org/10.20955/r.94.117-134

Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J.-B., Aali, A., Bluethgen, C., Pareek, A., Polacin, M., Reis, E. P., Seehofnerová, A., Rohatgi, N., Hosamani, P., Collins, W., Ahuja, N., Langlotz, C. P., Hom, J., Gatidis, S., Pauly, J., & Chaudhari, A. S. (2024). Adapted large language models can outperform medical experts in clinical text summarization. *Nature Medicine*, *30*(4), 1134–1142. https://doi.org/10.1038/s41591-024-02855-5

Venade, J., & Grilo, F. (2024). A Matter of Minutes: Unexpected FOMC Communication and Fed Credibility. *The B.E. Journal of Macroeconomics*. https://doi.org/10.1515/bejm-2024-0074

Xie, Q., Han, W., Zhang, X., Lai, Y., Peng, M., Lopez-Lira, A., & Huang, J. (2023). PIXIU: A Large Language Model, Instruction Data and Evaluation Benchmark for Finance. *Advances in Neural Information Processing Systems*, *36*.

Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., & Hashimoto, T. B. (2024). Benchmarking Large Language Models for News Summarization. *Transactions of the Association for Computational Linguistics*, *12*, 39–57. https://doi.org/10.1162/tacl_a_00632