# FedAgent: An Institution-Constrained Multi-Agent System Architecture for FOMC Meeting Simulation


## Abstract

Decisions of the Federal Open Market Committee (FOMC) regarding the federal funds rate play a central role in shaping financial market conditions and macroeconomic dynamics. However, forecasting the decisions is challenging due to the complicated committee-based nature. Most existing approaches map economic indicators or textual signals directly to policy actions, abstracting away from institutional deliberation. We propose FedAgent, an institution-constrained multi-agent framework that models FOMC interest rate decisions as a structured process of collective reasoning. The framework decomposes policymaking into evidence construction, discrete policy option formation, role-conditioned deliberation, and rule-based aggregation, while explicitly modeling role heterogeneity and asymmetric informational salience. Evaluated on scheduled FOMC meetings from 2020 to mid-2025, FedAgent achieves higher accuracy and greater prediction stability than strong machine learning and LLM-based baselines, while providing interpretable diagnostics of committee disagreement and consensus.


## 1 Introduction

Forecasting outcomes of committee-based decision-making is fundamentally different from standard prediction tasks. In institutions such as the Federal Open Market Committee (FOMC), policy actions emerge from structured deliberation among members with heterogeneous mandates, asymmetric information processing, and formal voting procedures, rather than from a single latent decision rule. Capturing such outcomes therefore requires modeling not only economic inputs and policy outputs, but also the institutional process through which collective reasoning unfolds.

Most existing approaches abstract away the institutional structure underlying monetary policy decisions. Both classical econometric models and modern machine learning methods map macroeconomic indicators directly to policy actions (Sarno et al., 2005; Yoon and Fan, 2024), bypassing the internal deliberation process of a committee. While recent advances in LLMs enable more realistic simulations of human reasoning (Park et al., 2023a), their direct application to FOMC-style decision-making remains limited. In particular, existing studies typically reduce policy outcomes to coarse directional categories (e.g., cut, hold, or hike), thereby overlooking disagreement over the magnitude of adjustment that is central to real-world FOMC deliberations(Seok et al., 2024; Kazinnik and Sinclair, 2025; Hou et al., 2025). As a result, these methods struggle to deliver stable behavior, interpretable disagreement among policymakers, and outputs that adhere to established institutional procedures.

We propose FedAgent, an institution-constrained multi-agent framework that reframes FOMC interest rate decisions as a structured collective reasoning problem. The framework explicitly separates evidence construction, policy option formation, role-conditioned deliberation, and rule-based aggregation, assigning LLM-based agents institutionally grounded roles with asymmetric informational salience. Crucially, the framework models policy decisions along both directional and intensive margins, allowing agents to express disagreement over the magnitude of rate adjustments rather than collapsing all dissent into coarse directional categories. By imposing a shared evidentiary baseline, a discretized and comparable policy space, and procedural constraints on deliberation and voting, the framework disciplines agent interaction and ensures that disagreement reflects institutional interpretation rather than stochastic dialogue dynamics. Empirical evaluation on U.S. monetary policy decisions from 2020 to mid-2025, which spans crisis easing, policy inertia, and rapid tightening periods, demonstrates that FedAgent achieves superior accuracy and greater stability relative to strong non-LLM and single agent baselines, with consistent performance across pronounced regime shifts.

Our contributions are threefold: First, we are the first to formulate FOMC interest rate forecasting as an institution-constrained collective reasoning problem and to propose a multi-agent framework that explicitly models deliberation, voting, and aggregation as endogenous components of policy determination. Second, we are the first to model role-induced heterogeneity within the committee, distinguishing Board Governors from Regional Bank Presidents, and to introduce a five-option discrete policy space that more closely reflects real world FOMC decision granularity than prior three-category schemes. Third, we demonstrate that the proposed framework delivers state-of-the-art performance across multiple monetary regimes with 83.53% accuracy and 93.27% Prediction Stability, outperforming a wide range of benchmarks such as FedSight and Machine Learning based models while providing interpretable diagnostics of internal consensus and dissent.

## 2 Related Work

### 2.1 Forecasting Federal Funds Target Rate

A large literature studies the prediction of Federal Funds Target Rate decisions using macroeconomic indicators, financial market expectations, and, more recently, textual information. Traditional econometric models and market-implied measures treat policy actions as direct functions of observed economic conditions, achieving reasonable performance in stable environments(Sarno et al., 2005). Machine learning approaches extend this line of work by introducing nonlinear models and high-dimensional predictors to improve short-horizon forecasts(Yoon and Fan, 2024; Chan-Lau et al., 2025). Parallel research applies text-as-data methods to central bank communications, including policy statements, meeting minutes, speeches, and regional reports, showing that textual signals contain incremental information about policy intentions and uncertainty(Hansen et al., 2018; Gössi et al., 2023). However, these approaches typically aggregate information into global representations and model policy outcomes as direct prediction targets, abstracting away the committee-based deliberation process through which heterogeneous information is interpreted, contested, and reconciled.

### 2.2 Large Language Models and Agent-Based Policy Modeling

Recent advances in large language models (LLMs) enable reasoning over unstructured text and heterogeneous inputs, motivating their application to economic forecasting, decision support, and simulation tasks (Wei et al., 2022; Bommasani, 2021). While structured prompting techniques such as chain-of-thought improve single-agent reasoning quality, they do not address settings where decisions are inherently collective (Wei et al., 2022). Emerging multi-agent LLM systems explore debate, collaboration, and consensus formation among interacting agents (Du et al., 2023; Park et al., 2023b), but often rely on unconstrained dialogue and task-agnostic interaction, limiting stability, reproducibility, and institutional alignment. In contrast to prior work, our approach explicitly models monetary policy decisions as an institution-constrained collective reasoning process, incorporating role heterogeneity, asymmetric information salience, a discretized policy action space, and formal aggregation to more faithfully reflect real world committee decision making.

More closely related literatures are few. MiniFed (Seok et al., 2024) organizes LLM agents into a five-stage workflow of discussion, persuasion, and voting, demonstrating that agentic interaction can reproduce realistic policy outcomes. FOMC in silico (Kazinnik and Sinclair, 2025) similarly models monetary policy as an agent-based process, emphasizing the potential of multi-agent systems to capture collective dynamics beyond direct prediction. FedSight AI (Hou et al., 2025) further explores structured multi-agent architectures for Federal Funds Target Rate forecasting, showing that coordinated agent reasoning can improve accuracy and interpretability. Our framework builds on these developments but departs in a fundamental way: rather than relying on unconstrained dialogue or workflow-centric coordination, we explicitly impose institutional constraints, role-induced asymmetries, a discretized and comparable policy action space, and rule-based aggregation. This design enables disagreement to arise from institutional interpretation rather than conversational dynamics, yielding a more faithful and stable simulation of FOMC deliberation.

## 3 System Architecture

### 3.1 Overview

We propose an institution-constrained multi-agent framework designed to simulate the decision-making processes of the FOMC. Our architecture (Figure 1) moves beyond unconstrained multi-agent dialogue by enforcing the structural and procedural realities of central banking: heterogeneous roles, asymmetric information access, and formalized voting protocols.

The system is organized into three modules, ensuring a unidirectional information flow. This design choice prevents information contamination, where agent deliberations might otherwise retroactively influence the factual baseline. The pipeline consists of the Data & Policy Module, Policymaker Module and Aggregation Module. The Data & Policy Module constructs the evidence base and discretizes the decision space. The Policymaker Module comprises 12 LLM-based agents with role-specific personas and information sets, and the Aggregation Module governs a multi-stage deliberation and voting process under a deterministic rule.

### 3.2 Data & Policy Module

The Data & Policy Module serves as the system's analytical foundation, responsible for transforming heterogeneous economic data into a structured, role-conditioned prompt space. To ensure the simulation's validity, we enforce a unidirectional information flow: upstream analysts construct a factual baseline that downstream policymaker agents cannot alter, thereby mitigating the risk of deliberative hallucination, where agents might otherwise invent or distort economic facts to suit a specific policy preference. The Module contains 4 agents: Macroeconomic Analyst, Market Analyst, Region Analyst and Policy Option Analyst.

Macroeconomic Analyst aggregates all available economic indicators (e.g. inflation, labor, growth) to produce a Standardized Macro Briefing prior to each FOMC meeting. The analyst will make sure all economic data be aligned to the pre-meeting blackout period to prevent look-ahead bias. Besides, the analyst is architecturally restricted to descriptive synthesis, explicitly prohibited from suggesting policy directions. This ensures that downstream disagreement among agents is a product of interpretative variance rather than inconsistent access to basic economic facts.

Market Analyst extracts market-implied expectations from interest rate futures and financial instruments before the meeting. This provides an external contrastive signal, representing the "market narrative." The market Analyst's purpose is not to guide or constrain agent decisions, but to provide an explicit contrastive signal against which agent reasoning can be evaluated, avoiding the implicit embedding of market priors within individual prompts.

A core innovation of our framework is the explicit modeling of institutional information asymmetry through the Region Analyst. The Region Analyst transforms district-specific Beige Book sections into strictly descriptive and traceable regional narrative digests, which are asymmetrically assigned to policymaker agents according to their institutional roles.

Reflecting the Federal Reserve's "7+5" voting structure, 7 Board Governors and 5 rotating Reserve Bank Presidents, our architecture captures the inherent tension between national economic oversight and regional economic conditions. Rather than treating the Beige Book as a monolithic document, the Region Analyst decomposes it into 12 district-level narrative digests that preserve localized economic signals often obscured in national aggregates.

Reserve Bank President agents assign high salience to the digest of their own district, while Governors rely primarily on national aggregates. This structured information asymmetry introduces interpretable heterogeneity into deliberation, stabilizes multi-agent interactions, and enables systematic analysis of how regional narratives shape committee-level outcomes.

The Policy Option Analyst serves as the interface between evidence construction and committee deliberation by organizing heterogeneous inputs into a shared and institutionally grounded action space for policymakers. It takes as input structured evidence summaries produced by upstream macroeconomic, market, and regional modules, and outputs a finite set of mutually comparable policy alternatives that constitute the exclusive choice set during the voting stage.

Formally, the analyst constructs a discrete policy menu

$$\mathcal{O} = \{o_1, o_2, o_3, o_4, o_5\},$$

where each option $o_k \in \mathcal{O}$ corresponds to a predefined basis-point change

$$\Delta r_k \in \{-50, -25, 0, +25, +50\}.$$

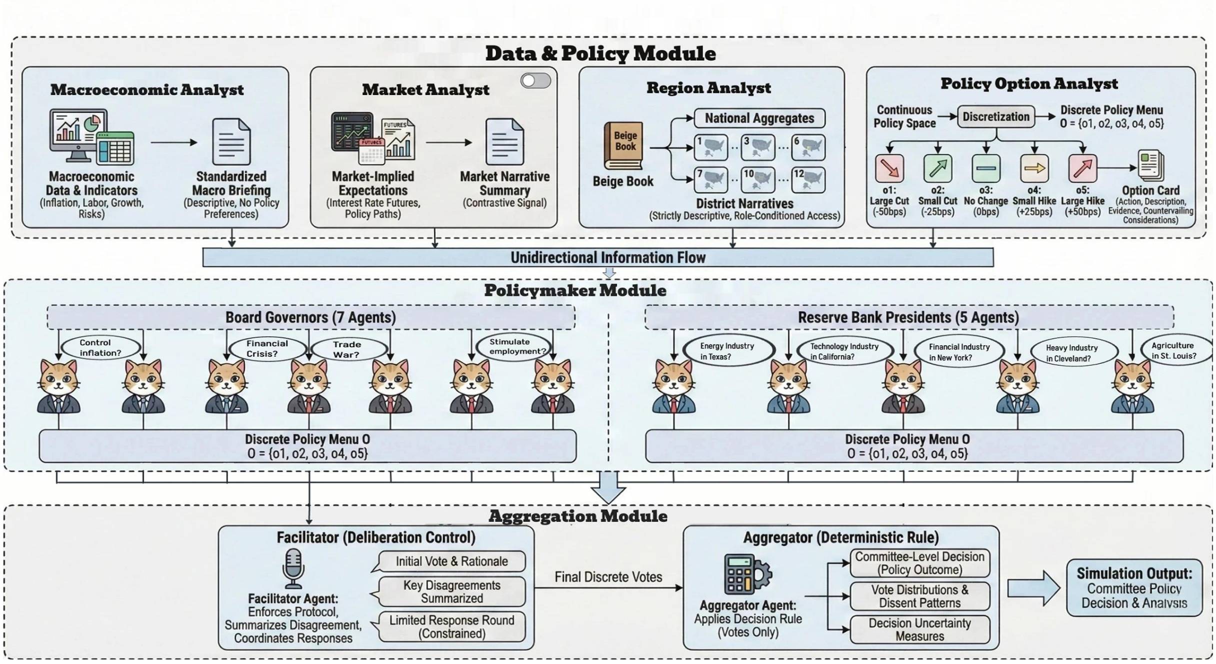


Figure 1: Workflow of FedAgent multi-agent system.

Each option is represented as a structured policy card consisting of (i) the numerical rate adjustment $\Delta r_k$, (ii) a neutral description of the policy stance, (iii) supporting evidence drawn from the inputs, and (iv) countervailing considerations reflecting key risks or uncertainties.

Compared to coarse the conventional dovish–neutral–hawkish 3–option schemes, this design enables finer-grained expression of policy stances while remaining tractable, and more closely reflects real-world FOMC deliberations, particularly during periods of acutual financial stress or economic downturns when larger and asymmetric policy adjustments become salient.

### 3.3 Policymaker Module

The Policymaker Module represents the deliberative core of the system, translating a fixed evidentiary baseline into institutionally grounded policy judgments. It consists of twelve LLM-based agents whose composition mirrors the voting structure of the Federal Open Market Committee, thereby embedding institutional heterogeneity directly into the decision process.

Agents are not treated as interchangeable reasoning units. Instead, each operates under a role-conditioned professional persona reflecting its formal mandate. Board Governor agents prioritize system-wide objectives such as inflation control, labor market conditions, financial stability, and global spillovers, while Reserve Bank President agents assign elevated salience to region-specific economic dynamics tied to their respective districts. Although all agents consume the same standardized evidence, these role-based filters govern how information is internally weighted and interpreted, allowing disagreement to arise endogenously from institutional mandates rather than from asymmetric data access. All agents are strictly bounded by the shared discrete policy menu $\mathcal{O}$. This constraint shifts reasoning away from open-ended generation toward comparative evaluation among admissible actions. Agents must therefore justify why a particular option dominates its alternatives given their institutional objectives and the observed evidence, rather than proposing novel or unconstrained policy paths.In each simulation round, agents select a policy option $o_k \in \mathcal{O}$, provide a structured rationale linking evidence to mandate, and report a calibrated confidence score. Together, these elements define a complete decision state, enabling downstream aggregation to distinguish substantive disagreement from weakly held preferences.

### 3.4 Aggregation Module

The Aggregation Module synthesizes heterogeneous agent judgments into a single committee decision through a deterministic and institutionally grounded protocol. Its design explicitly separates deliberation from decision-locking, ensuring that collective reasoning can evolve without compromising the stability of the final outcome.

The process begins with a facilitated deliberation stage. A dedicated Facilitator Agent reviews the initial distribution of votes and rationales and identifies the primary dimensions of disagreement, such as conflicting assessments of inflation persistence or asymmetric regional risks. Rather than amplifying the majority position, the Facilitator distills these disagreements into a concise set of contested claims. Agents then engage in a tightly bounded deliberation round, limited in both length and scope, during which responses must directly address the identified points using evidence-based arguments. This structure mirrors the focused and adversarial nature of real-world FOMC discussions while mitigating semantic drift and conversational redundancy.

Following deliberation, the system transitions to a hard aggregation stage. The Aggregator Agent applies a fixed mathematical rule to the terminal set of discrete votes, producing the final policy decision. By decoupling the outcome from narrative persuasion, the framework ensures that the decision reflects the committee's resolved preference structure rather than rhetorical variation across agents.

In addition to the final action, the module outputs vote distributions and confidence-weighted dispersion measures, providing transparent diagnostics of internal consensus and dissent.

# 4 Empirical Results

## 4.1 Experimental Settings

We simulated all scheduled FOMC meetings from January 2020 to June 2025, excluding unscheduled or emergency interventions[1], yielding 43 meeting instances. Each meeting was aligned with its real-world counterpart and simulated independently under a fixed system architecture. To ensure temporal validity and mitigate hallucination induced by information leakage, all structured indicators and market signals were truncated to information available prior to the meeting, while qualitative inputs such as the Beige Book were restricted to the most recent pre-meeting release. For each meeting, the full pipeline, including evidence construction, policy option formation, agent voting constrained deliberation and aggregation, was executed repeatedly. To account for stochasticity in LLM generation, each meeting was simulated 50 times with identical inputs, and results were averaged across runs. All experiments were conducted on our custom multi-agent simulation platform, which interfaces directly with the OpenAI GPT-4o to instantiate role conditioned agents and enforce institutional constraints. Approximately 80 million tokens are used across all agent-related experiments and comparisons[2].

## 4.2 Baseline Methods

We benchmark our framework against a compact set of baselines spanning traditional machine learning, neural networks, and LLM-based systems. We adopt LightGBM(LGBM) using the same economic inputs, and further include an ordinal random forest model(ORF) that explicitly accounts for the ordered nature of policy actions, following the strongest non-LLM baseline reported in FedSight. We also evaluate a standard multilayer perceptron (MLP) to assess whether generic neural networks suffice for this task in the absence of language-based reasoning. In addition, we compare against several LLM-based baselines, including a single-agent LLM (SAgent) without committee structure, the original FedSight multi-agent system, and FedSight augmented with Chain-of-Draft (CoD), which constitutes the strongest published LLM-based baseline for FOMC decision prediction. All baselines operate on the same discrete policy option space where applicable, ensuring that performance differences reflect modeling assumptions rather than evaluation artifacts.

## 4.3 Performance Metrics

We evaluate all models using a set of metrics that jointly assess predictive accuracy, robustness, and, where applicable, agent-level reliability and institutional behavior.

- **Accuracy**

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{\hat{y}_i = y_i\},$$

measuring the proportion of FOMC meetings for which the predicted policy action $\hat{y}_i$ exactly matches the realized decision $y_i$, where $i$ indexes meetings and actions are drawn from the discrete option set.

[1] We excluded Meeting in 2020 March and 2020 August because they are either unscheduled or notation vote.

[2] For tokens used, FedSight and FedSight(CoD) use about 30 million. MiniFed use about 17 million. Our agent framework use about 32 million tokens

- **Mean Absolute Error (MAE)**

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |\hat{y}_i - y_i|,$$

where policy options are mapped to numerical values $\{-50, -25, 0, +25, +50\}$ (in basis points). MAE distinguishes near misses from larger directional errors and is applicable to all models.

- **Prediction Stability (PreStability)**

$$\mathrm{PreStability} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{\hat{y}_i^{(j)} = \hat{y}_i^{\mathrm{mode}}\},$$

quantifying the consistency of meeting-level predictions across repeated simulations with identical inputs, where $j$ indexes simulation runs and $\hat{y}_i^{\mathrm{mode}}$ denotes the modal prediction. This metric is reported for stochastic models.

- **Agent Accuracy**

$$\mathrm{AgentAcc} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{K}\sum_{k=1}^{K} \mathbf{1}\{\hat{v}_{i,k} = y_i\},$$

representing the average fraction of individual agent votes $\hat{v}_{i,k}$ that align with the realized committee decision, where $k$ indexes policymaker agents. This metric is defined only for agent-based models.

- **Voting Stability**

$$\mathrm{VoteStability} = \mathrm{avg}_{i,j,k}\, \mathbf{1}\{\hat{v}_{i,k}^{(j)} = \hat{v}_{i,k}^{\mathrm{mode}}\},$$

measuring the consistency of agent-level votes across repeated simulations, capturing the robustness of deliberative outcomes.

- **Dissent Rate**

$$\mathrm{Dissent} = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K} \mathbf{1}\{\hat{v}_{i,k} \neq \hat{y}_i\},$$

quantifying the frequency with which individual agents dissent from the committee-level decision, providing a measure of institutional plausibility and internal disagreement.

### 4.4 Results

Table 1 summarizes our empirical results. Among non-LLM baselines, LGBM and MLP achieve moderate accuracy, while ORF performs best, consistent with prior evidence that exploiting the ordered structure of policy actions is beneficial. Neural networks provide limited gains over strong tree-based models.

Single agent achieves accuracy comparable to ORF, indicating that language-based reasoning can partially substitute for handcrafted features. However, models that explicitly simulate committee deliberation consistently outperform single-agent approaches. FedSight substantially improves over the single-agent, and the addition of Chain-of-Draft reasoning yields further gains in both accuracy and stability.

Our FedAgent achieves the strongest overall performance, with the highest accuracy (83.53%), the lowest MAE (0.1107), and the greatest prediction stability (93.27%). These improvements indicate that structured deliberation and institutional constraints contribute meaningfully beyond model capacity alone[3].

Table 2 reports agent-level accuracy, voting stability, and dissent rates. Relative to FedSight and its CoD variant, FedAgent exhibits higher agent accuracy and more consistent voting behavior across simulations. Importantly, dissent rates remain nontrivial (12.93%), suggesting that performance gains are not driven by artificial consensus but arise from disciplined deliberation under asymmetric information and procedural constraints.

### 4.5 Ablation Study

Table 3 reports module-level ablations. Removing the market module leads to a sharp drop in accuracy and stability, indicating its role in anchoring deliberation rather than inflating accuracy alone. Disabling the policy module causes the largest stability loss, highlighting the importance of explicit option framing. Removing the region module mildly reduces accuracy but substantially increases dissent, reflecting the loss of structured heterogeneity. Replacing the committee with a single agent or removing aggregation both significantly degrade performance and robustness. Overall, each module contributes complementarily, and the full system is required for stable, institutionally grounded delib-

[3]We also compare FedAgent with MiniFed (Seok et al., 2024). The results are shown in Appendix B.

| Metrics | LGBM | ORF | MLP | SinAgent | FedSight | FedSight(CoD) | FedAgent |
|---|---|---|---|---|---|---|---|
| Accuracy | 66.17% | 69.56% | 68.42% | 71.81% | 78.34% | 81.84% | 83.53% |
| MAE | 0.3124 | 0.2207 | 0.2715 | 0.2495 | 0.1648 | 0.1325 | 0.1107 |
| PreStability | N/A | N/A | N/A | 81.16% | 84.35% | 90.62% | 93.27% |

Table 1: Predictive performance across baselines

| Metrics | FedSight | FedSight(CoD) | FedAgent |
|---|---|---|---|
| Agent Accuracy | 74.56% | 77.41% | 80.74% |
| Vote Stability | 82.36% | 85.19% | 87.35% |
| Dissent Rate | 18.16% | 14.02% | 12.93% |

Table 2: Predictive performance across baselines

eration.

### 4.6 Regime-Specific Analysis (2020–2025)

We further evaluate robustness under non-stationary policy environments by examining performance across distinct monetary regimes from 2020 to 2025. We partition the sample into three regimes based on realized policy dynamics: (i) emergency easing during the COVID-19 crisis (2020), (ii) an extended hold period with near-zero rates (2021), (iii) the rapid tightening cycle (2022–2023)

As shown in Table 4, the model maintains stable performance in different regimes. Accuracy is lowest during the 2020 emergency easing period (81.56%), reflecting heightened uncertainty, but stability remains high (90.74%), indicating controlled reasoning rather than erratic behavior. Performance peaks during the 2021 hold regime, with the highest accuracy (88.37%) and stability (95.35%), without collapsing into trivial majority predictions. The tightening phase of 2022–2023 presents the most challenging dynamics; although accuracy declines modestly, both MAE and stability remain well controlled. Overall, the results demonstrate that the proposed framework delivers consistent accuracy and stable deliberation across heterogeneous policy regimes, supporting its robustness under severe non-stationarity.

## 5 Conclusion

We propose FedAgent, an institution-aware multi-agent framework for forecasting committee-based policy decisions, using the FOMC as a case study. By explicitly modeling role heterogeneity, asymmetric information, constrained deliberation, and formal aggregation, the system moves beyond black-box prediction toward process-grounded collective reasoning. Empirical results from 2020 to 2025 show that the proposed framework consistently outperforms strong non-LLM baselines and single-agent LLMs, achieving higher accuracy (83.53%) ,higher prediction stability(93.27%), and lower dissent rate (12.93%). Module-level ablations and regime-specific analyses demonstrate that these gains arise from complementary architectural components rather than heuristics or overfitting to specific periods. Agent-level diagnostics further show realistic dissent and robust voting behavior, indicating disciplined deliberation rather than artificial consensus. Together, these findings highlight the value of institutionally grounded multi-agent design for modeling complex collective decisions under non-stationarity, with broader implications for applying language-based agents to real-world decision-making systems.

## 6 Limitations

First, the empirical evaluation is constrained by the limited number of FOMC meetings, especially in regime-specific analyses. Although the sample spans multiple policy regimes from 2020 to 2025, results should be interpreted as indicative rather than statistically definitive. Second, the framework relies on curated representations of macroeconomic indicators, market expectations, and regional narratives. While designed to reflect meeting-available information, preprocessing and summarization may abstract away nuances present in real-world deliberations. Third, large language models introduce inherent stochasticity and sensitivity to prompts and model configurations, limiting direct transferability across architectures or future model versions. Fourth, although institutional roles and procedures are explicitly modeled, informal coordination, interpersonal influence, and strategic communication

| Setting | Accuracy | MAE | PreStability | AgAccuracy | VotStability | Dissent |
|---|---|---|---|---|---|---|
| FedAgent | 83.53% | 0.1107 | 93.27% | 80.74% | 87.35% | 12.93% |
| w/o Market Analyst | 72.65% | 0.2847 | 80.31% | 70.34% | 81.05% | 22.19% |
| w/o Policy Option | 74.32% | 0.2475 | 83.24% | 72.85% | 83.17% | 18.35% |
| w/o Region Analyst | 77.18% | 0.2083 | 88.26% | 76.35% | 89.03% | 25.27% |
| w/o Policymaker | 71.22% | 0.3059 | 85.81% | 68.34% | 83.84% | 27.53% |
| w/o Aggregation | 71.53% | 0.2962 | 82.81% | 71.34% | 82.84% | 26.78% |

Table 3: Ablation Study

| Regimes | Accuracy | MAE | Stability |
|---|---|---|---|
| 2020 Emergency Easing | 81.56% | 0.1734 | 90.74% |
| 2021 Policy Hold | 88.37% | 0.0961 | 95.35% |
| 2022–2023 Tightening | 80.16% | 0.2015 | 91.93% |

Table 4: Regime-specific performance across monetary policy phases.

are outside the scope of the current design. Finally, the study focuses on a single policy institution; extending the framework to other committee-based decision settings remains an important direction for future work.

## References


Rishi Bommasani. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Jorge A Chan-Lau, Toan Long Quach, and Aruhan Rui Shi. 2025. Forecasting federal funds rates with ai: Lstm, gru, and large language model approaches. *GRU, and Large Language Model Approaches (July 01, 2025)*.

Yilun Du, Shuang Li, Antonio Torralba, Joshua B Tenenbaum, and Igor Mordatch. 2023. Improving factuality and reasoning in language models through multiagent debate. In *Forty-first International Conference on Machine Learning*.

Sandro Gössi, Ziwei Chen, Wonseong Kim, Bernhard Bermeitinger, and Siegfried Handschuh. 2023. Finbert-fomc: Fine-tuned finbert model with sentiment focus method for enhancing sentiment analysis of fomc minutes. In *Proceedings of the Fourth ACM International Conference on AI in Finance*, pages 357–364.

Stephen Hansen, Michael McMahon, and Andrea Prat. 2018. Transparency and deliberation within the fomc: A computational linguistics approach. *The Quarterly Journal of Economics*, 133(2):801–870.

Yuhan Hou, Tianji Rao, Jeremy Tan, Adler Viton, Xiyue Zhang, David Ye, Abhishek Kodi, Sanjana Dulam, Aditya Paul, and Yikai Feng. 2025. Fedsight ai: Multi-agent system architecture for federal funds target rate prediction. *arXiv preprint arXiv:2512.15728*.

Sophia Kazinnik and Tara M Sinclair. 2025. Fomc in silico: A multi-agent system for monetary policy decision modeling. *Available at SSRN 5424097*.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023a. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023b. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22.

Lucio Sarno, Daniel L Thornton, and Giorgio Valente. 2005. Federal funds rate prediction. *Journal of Money, Credit and Banking*, pages 449–471.

Sungil Seok, Shuide Wen, Qiyuan Yang, Juan Feng, and Wenming Yang. 2024. Minifed: Integrating llm-based agentic-workflow for simulating fomc meeting. *arXiv preprint arXiv:2410.18012*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Jungyeon Yoon and Juanjuan Fan. 2024. Forecasting the direction of the fed's monetary policy decisions using random forest. *Journal of Forecasting*, 43(7):2848–2859.

## A Macroeconomic Variables

Table 5 shows all macroeconomic variables used for all machine learning models and LLM-Based models.

## B Compare MiniFed with FedAgent

Consistent with (Hou et al., 2025), we note that the experimental evaluation in (Seok et al., 2024) is conducted exclusively on the 2018 FOMC meetings. Moreover, the authors do not publicly release the full set of prompts and system configurations required for exact replication. To ensure a fair and transparent comparison, we therefore adopt their reported 2018 results as given and directly benchmark our model on the same 2018 test set.

## C Selected Prompts

| Variable | Description |
|---|---|
| **Inflation Variables** | |
| PCE Inflation(Monthly) | Change in PCE Price Index |
| Core PCE Inflation(Monthly) | Core PCE inflation excluding food and energy |
| CPI Inflation(Monthly) | Monthly change in Consumer Price Index |
| Core CPI Inflation(Monthly) | CPI inflation excluding food and energy |
| Breakeven Inflation(Daily) | 5-year TIPS breakeven inflation rate |
| **Labor Market Variables** | |
| Unemployment Rate(U-3) | Civilian unemployment rate |
| Unemployment Rate (U-6) | Broad unemployment and underemployment rate |
| Labor Force Participation | Labor force participation rate |
| Nonfarm Payroll Growth(Monthly) | Change in nonfarm payroll employment |
| Job Openings Rate(Monthly) | Job openings rate from JOLTS |
| Job Openings / Unemployed (Monthly) | Measure of labor market tightness |
| Quit Rate (Monthly) | Voluntary job separation rate |
| **Economic Activity Variables** | |
| GDP Growth | High-frequency real-time GDP growth estimate |
| Industrial Production | Growth rate of industrial production index |
| Retail Sales | Growth rate of real retail sales |
| Personal Consumption (Monthly) | Growth rate of personal consumption expenditures |
| ISM Manufacturing PMI (Monthly) | Manufacturing Purchasing Managers Index |
| ISM Services PMI (Monthly) | Services sector Purchasing Managers Index |
| **Monetary Policy Variables** | |
| Federal Funds Target Rate | Upper bound of the federal funds target range |
| Previous FFTR | Target rate prior to current meeting |
| Previous Rate Change | Basis point change at prior meeting |
| 3-Month Treasury Yield (Daily) | 3-month Treasury bill yield |
| 6-Month Treasury Yield(Daily) | 6-month Treasury bill yield |
| 2-Year Treasury Yield (Daily) | 2-year Treasury bill yield |
| 10-Year Treasury Yield(Daily) | 10-year Treasury yield |
| Yield Curve Slope (10Y–3M) | Term spread between 10-year and 3-month yields |
| Fed Total Assets (Weekly) | Total assets on Federal Reserve balance sheet |
| Reserve Balances (Weekly) | Reserve balances held by depository institutions |
| **Financial Conditions Variable** | |
| VIX Index (Daily) | Equity market volatility index |
| Credit Spread (BAA–10Y) (Daily) | Corporate bond spread over Treasuries |
| Equity Market Return (Daily) | S&P 500 daily return |
| Equity Market Volatility (Daily) | Realized equity market volatility |

Table 5: Macroeconomic Variables

| FOMC | Actual | MiniFed | FedAgent |
|---|---|---|---|
| Jan 2018 | 0.00% | 0.25% | 0.00% |
| Mar 2018 | 0.25% | 0.25% | 0.25% |
| May 2018 | 0.00% | 0.00% | 0.00% |
| Jun 2018 | 0.25% | 0.25% | 0.25% |
| Aug 2018 | 0.00% | 0.00% | 0.00% |
| Sep 2018 | 0.25% | 0.00% | 0.25% |
| Nov 2018 | 0.00% | 0.00% | 0.00% |
| Dec 2018 | 0.25% | 0.25% | 0.25% |

Table 6: Compare MiniFed with FedAgent

| Function | Prompts |
|---|---|
| Macroeconomic Analyst | You will act as a Macroeconomic Analyst for the Federal Open Market Committee. Your task is to summarize the current macroeconomic environment prior to the meeting. Focus exclusively on describing inflation dynamics, labor market conditions, and real economic activity using the provided indicators. All information must be restricted to data available before the blackout period. Do not provide policy recommendations or normative judgments. |
| Market Analyst | You will act as a Market Analyst summarizing market-implied expectations. Based on interest rate futures, yield curves, and financial conditions, describe how financial markets appear to price near-term monetary policy. Your role is purely descriptive and contrastive: do not endorse or recommend any policy action. Use only market information available prior to the meeting. |
| Regional Analyst | You will act as a Regional Analyst processing the Beige Book. Carefully read the district-specific Beige Book section and produce a factual summary of regional economic conditions, including activity, labor, prices, and sentiment. Do not generalize beyond the text or infer national implications. Your output should remain strictly descriptive and traceable to the source material. |
| Policy Option Analyst | You will act as a Policy Option Analyst. Based on the macroeconomic briefing, market summary, and regional digests, organize the evidence into a set of five discrete monetary policy options: a large cut, a small cut, no change, a small hike, and a large hike. Each option should be described neutrally, with supporting considerations and countervailing risks. Do not indicate which option is preferred. |

Table 7: Selected Prompts for Modules

| Function | Prompts |
|---|---|
| Character Definition | You will play the role of Jerome H. Powell, Chair of the Board of Governors of the Federal Reserve System, participating in the current Federal Open Market Committee meeting. In all deliberations, you should respond faithfully as the Federal Reserve Chair, reflecting both institutional responsibilities and personal leadership style. |
| Background | You were first appointed to the Federal Reserve Board by President Barack Obama and were subsequently appointed as Chair by President Donald Trump, later reappointed by President Joseph R. Biden Jr. You are a non-partisan central banker without formal party affiliation. You are currently serving an active Chair term with remaining N-year tenure, and are therefore concerned more with policy credibility, institutional continuity, and long-run macroeconomic stability. |
| Institutional Mandate | As Chair, your primary mandate is to promote maximum employment and price stability, while preserving financial stability and the credibility of the Federal Reserve as an independent institution. You should balance short-term macroeconomic developments against medium-term inflation expectations and systemic risks. |
| Information Access | You base your judgment primarily on national-level information, including: (1) standardized macroeconomic indicators from the macroeconomic analyst, (2) market-implied expectations and financial conditions from the market analyst, (3) high-level summaries of regional conditions from the region market, (4) the complete discrete policy menu constructed by the Policy Option Analyst. You do not rely on idiosyncratic district-level anecdotes unless they signal broader systemic trends. |
| Leadership Style | Your leadership style is pragmatic, consensus-oriented, and institutionally cautious. You value listening to diverse viewpoints within the Committee and aim to guide deliberation toward a coherent and credible collective outcome. You avoid ideological rigidity and emphasize clear communication and risk management. |
| Historical Policy Tendency | Historically, your policy stance is best characterized as data-dependent and risk-balanced. You are neither persistently hawkish nor dovish: you are willing to support accommodative policy during crises or severe downside risks, but you place strong emphasis on anchoring inflation expectations when inflation pressures become persistent. You are especially attentive to asymmetric risks and nonlinearities. |
| Deliberation Behavior | During discussions, articulate your reasoning clearly and professionally, referencing national macroeconomic conditions and financial stability considerations. You may acknowledge uncertainty and trade-offs, and you should engage constructively with opposing views in order to foster convergence within the Committee. |
| Policy Decision | Based on the available evidence and the discrete policy alternatives presented, select the single option you judge to be most appropriate at this meeting. Your final vote should reflect a careful balance between inflation risks, labor market conditions, and the Federal Reserve's institutional credibility. Provide a concise economic justification and identify key uncertainties influencing your confidence. |

Table 8: Selected Prompts for the Policymaker Agent (Governor): Chair Jerome H. Powell