13th ECB Conference on Forecasting Techniques

FOMC *In Silico*: A Multi-Agent System for Monetary Policy Decision Modeling

Sophia Kazinnik & Tara M. Sinclair

March 23, 2026

Why This Matters for Central Banks

Central banks do not make decisions through a mechanical reaction function alone:

- They make decisions through **institutions**: committees, agenda setting, persuasion, dissent, and communication.
- Traditional models are useful for the macro side of the problem.
- But they usually abstract from the **human layer** of policy:
  - who influences whom,
  - how members interpret the same data differently,
  - how consensus forms,
  - and how political or institutional pressure changes behavior.

This is where **AI agents** may be useful: not as replacements for policymakers, but as tools for modeling the policy process itself.

Why AI Agents?

Growing body of literature shows that LLMs produce responses consistent with both economic theory and documented patterns of human behavior:

- Horton, 2023 (Economics)
- Dillion et al., 2023; Kosinski, 2023 (Psychology)
- Bail, 2024; Ashokkumar et al., 2024 (Sociology)
- Aubin Le Quéré et al., 2024; Park et al., 2024 (Computer Science)

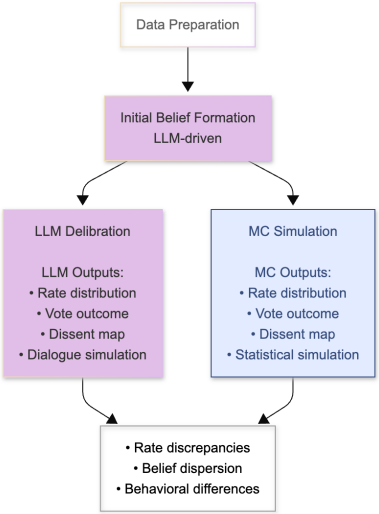
Could we then:

- use **AI agents** as a structured way to encode:
  - member priors, heterogeneity, speeches;
  - institutional norms and agenda power
- ..and run policy *in silico*?

A dual-track simulation of the FOMC

- **Phase 1:** build the macro snapshot, members, and member-specific priors.
- **Phase 2:** let the synthetic committee deliberate and form consensus.
- **Phase 3:** compare the final vote to a benchmark using the same inputs but no deliberation.

The benchmark gives a rational baseline; the LLM track shows how communication and institutional frictions shift outcomes.



Phase One:
Data, Agents, & Initial Beliefs

Macro data

Recent financial news

Synthetic Beige Book

Previous Announcement

Data Update

```
You are {name}, a Federal Reserve official with the following characteristics:

Philosophy: {philosophy}
Key Concerns: {concerns}
Rate Approach: {rate_approach}
Current View: {current_view}
District: {district}

Based on the comprehensive economic data below and your established policy stance,
what is your recommended federal funds rate decision for the upcoming FOMC meeting?

COMPREHENSIVE ECONOMIC DATA AND CONTEXT:
{comprehensive_data}

Consider:
1. The current fed funds rate and recent FOMC decisions
2. Key economic indicators (inflation, employment, growth)
3. Your district's specific economic conditions
4. Market expectations and recent economic news
5. Your historical policy stance and priorities
```

Biography

Known policy stance

Recent speeches

FOMC Member #N

Hold rate (0 bp change)
confidence: 0.85

Repeat for each voting member (× 12)

Data Inputs: Macro Data

- Snapshot for a single meeting (here, July 2025):
  - Inflation (headline/core), labor market, output growth, financial conditions, etc.
  - Can be updated continuously, if needed (e.g., daily basis)
- Automatically accessing FRED, BLS, news via API

Data Inputs: Macro Dashboard

Table 1: Key U.S. Economic Indicators – Snapshot as of July 31st, 2025

Indicator	Latest	YoY	Trend	Source
Growth & Output				
Real GDP (2025 Q1)	$23,685.3 bn	2.0 %	↑	FRED
Industrial Production (Index)	104.0	0.7 %	↑	FRED
Retail Sales (monthly)	$720.1 bn	3.9 %	↑	FRED
Housing Starts	1.3 m	−0.5 %	↓	FRED
Inflation & Prices				
CPI (All Items)	321.5	2.7 %	↑	FRED
Core CPI	327.6	2.9 %	↑	FRED
PCE (All Items)	126.6	2.6 %	↑	FRED
Core PCE	125.9	2.8 %	↑	FRED
WTI Crude Oil ($/bbl)	$67.8	−12.7 %	→	FRED
Labor Market				
Unemployment Rate	4.1 %	0.0 pp	↑	BLS
Nonfarm Payrolls	159.7 m	1.2 %	↑	BLS
Labor Force Participation	62.3 %	−0.5 pp	↓	BLS
Average Hourly Earnings	$31.2	3.9 %	↑	BLS
Employment-Population Ratio	59.7 %	−0.5 pp	↓	BLS
Financial Conditions				
Fed Funds Rate	4.3 %	−75.0 bp	→	FRED
10-yr Treasury Yield	4.4 %	16.0 bp	↓	FRED
2-yr Treasury Yield	3.9 %	−18.0 bp	↓	FRED
S&P 500 Index	6,362.9	19.1 %	↑	FRED
U.S. Dollar Index (DXY)	120.4	−2.8 %	↓	FRED
VIX Volatility Index	15.5	−25.3 %	↓	FRED

Data Inputs: Synthetic Beige Book

- The Beige Book contains narrative summaries of regional economic conditions

- We construct a *synthetic*, real-time version:
  - Start from the most recent official Beige Book
  - Augment with web intelligence (news, gov and industry sources) via a research agent (`GPTResearcher`)

Data Inputs: Synthetic Beige Book

- Implementation:
  - `GPTResearcher` (open deep research agent) with 12 "digital analysts", one per district
  - Automatically querying government agencies, trade groups, regional media, and research centers over a 2-week window
  - LLM synthesizes district reports into a structured comparison table (growth category, trend arrows, consistency vs. official Beige Book)
- Why it matters:
  - Provides qualitatively rich but structured input
  - Mirrors the narrative information human FOMC members actually see

Data Inputs: Synthetic Beige Book

Table 2: District Economic Conditions:
Official *Beige Book* vs. Synthetic Version

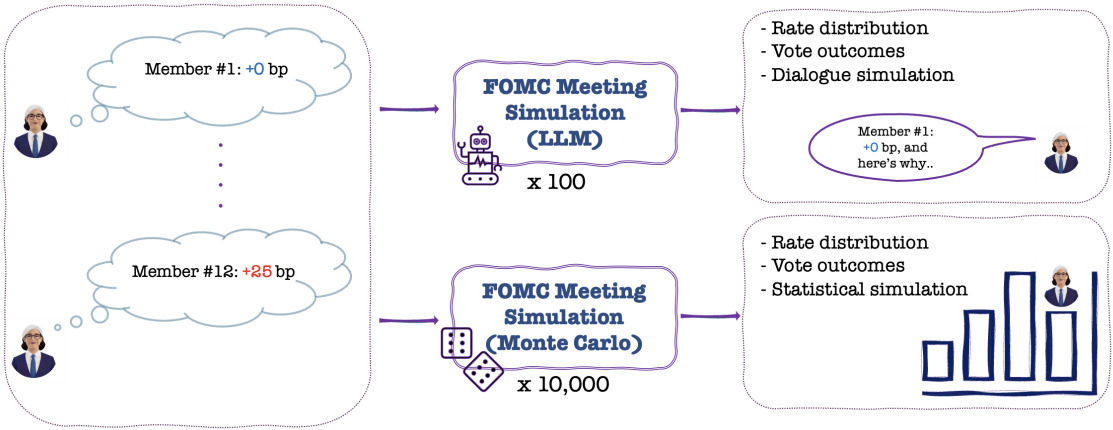
District	Official	*sim*Current	Consistency	Trend
Boston (1st)	Modest	Moderate	Consistent	↑
New York (2nd)	Modest	Moderate	Consistent	↑
Philadelphia (3rd)	Moderate	Mixed	Shifted	↓
Cleveland (4th)	Modest	Mixed	Shifted	↓
Richmond (5th)	Moderate	Moderate	Consistent	→
Atlanta (6th)	Moderate	Slowing	Shifted	↓
Chicago (7th)	Mixed	Moderate	Consistent	↑
St. Louis (8th)	Modest	Moderate	Consistent	↑
Minneapolis (9th)	Moderate	Moderate	Consistent	→
Kansas City (10th)	Modest	Moderate	Consistent	↑
Dallas (11th)	Robust	Slowing	Shifted	↓
San Francisco (12th)	Moderate	Robust	Consistent	↑

Growth scale: Robust, Moderate, Modest, Mixed, Slowing.
Consistency: Consistent = assessments differ by ≤ 1 step; Shifted = swing of ≥ 2 steps.
Trend arrows: ↑ improving, → stable, ↓ weakening.

FOMC Personas: From Text to Priors

- Next, for each voting member, we build a *synthetic persona* from:
  - Biography and role (Governor/Board Member/President), career history, political affiliation
  - Recent speeches with LLM stance classification (HAWKISH / DOVISH / DATA-DEPENDENT):
    - Temporal weighting: most recent speech $\times 2.5$, second $\times 1.3$, ..
- Prompted persona fields (natural language $\rightarrow$ JSON):
  - `philosophy`, `concerns`, `rate_approach`, `communication_style`, `current_view`, `biographical_influences`

Phase Two:
Deliberations and Voting


Member #1: +0 bp
Member #12: +25 bp
FOMC Meeting Simulation (LLM)
x 100
FOMC Meeting Simulation (Monte Carlo)
x 10,000
- Rate distribution
- Vote outcomes
- Dialogue simulation
Member #1: +0 bp, and here's why..
- Rate distribution
- Vote outcomes
- Statistical simulation
Step 1: Load Initial Opinions
Step 2: Debate & Update
Step 3: Final Vote & Outcomes

LLM Deliberation: Meeting Flow

1. **Initialization**
   - Load personas from Part 1 (with $\mu_i^{(0)}, \sigma_i^{(0)}$)
   - Build system prompt with all 12 member descriptions
2. **Deliberation rounds** (typically 2–3)
   - Chair opens: "Let's begin our policy discussion..."
   - Each member speaks (LLM generates in-character response + JSON signal)
   - Signal extraction: $\mu_i^{(t+1)} \leftarrow$ parsed `signal_rate` from LLM
   - Transcript persists across rounds; later speakers can reference earlier points
3. **Voting**: Members produce final rate signals (their preferred rate and uncertainty), and then the model aggregates them. The final rate is a weighted average of the Chair's agenda rate and the committee median (stylized RRM 2010).
4. **Meeting outcome**: final rate, dissenter list, full transcript, FOMC announcement

MC Benchmark (stylized RRM)

1. **Inputs (same as LLM)**
   - Start from member priors from Phase 1: $(\mu_i^{(0)}, \sigma_i^{(0)})$
   - (No dialogue; no transcript)
2. **Monte Carlo draw** (repeat $M = 10{,}000$ times)
   - **Preference realization:** draw a realized ideal point around $\mu_i^{(0)}$
   - **Private signal:** draw a noisy signal s_i for each member
   - **Bayesian update:** combine prior + signal $\Rightarrow$ posterior $(\mu_i^{\text{post}}, \sigma_i^{\text{post}})$
3. **Chair agenda setting**
   - Propose agenda rate r_c using a deterministic rule (median) + bias Δ
4. **Voting + outcome**
   - Members vote YES/NO based on proximity to μ_i^{post} and dissent costs
5. **Outputs**
   - Distribution over r_f (mean, dispersion, tails) and dissent counts

Voting Mechanism (stylized RRM 2010)

1. **Chair proposes** agenda rate r_c based on strategy:

$$r_c = \begin{cases} \text{median}\{\mu_i^{\text{post}}\} + \Delta & \text{(median strategy)} \\ \text{mean}\{\mu_i^{\text{post}}\} + \Delta & \text{(mean strategy)} \\ r_{-1} & \text{(status quo strategy)} \end{cases}$$

   where Δ is chair bias

2. **Members vote to approve** if close enough to their preference:

$$\text{Approve} \iff \left| r_c - \mu_i^{\text{post}} \right| \leq \varepsilon \cdot \kappa$$

   $\varepsilon = 0.20$ (20bp tolerance window), κ = dissent cost multiplier

3. **Aggregation rule:**

$$r_f = w_a \cdot r_c + (1 - w_a) \cdot \text{median}\left(\mu_i^{\text{post}}\right)$$

Three Scenarios

Three scenarios, one shared information set

The macro backdrop is fixed; only institutional conditions change.

Baseline

- Late July 2025 data
- Committee starts mildly hawkish
- Both tracks remain inside the target range

Political pressure

- Chair authority weakened
- Agenda tilts more dovish
- Disagreement becomes more likely

Payroll revision

- Softer labor signal added
- Higher uncertainty around the outlook
- Disagreement becomes more likely

Baseline: small wedge, broad consensus

- Both tracks choose rates inside the **4.25–4.50%** target band.
- Mean outcome:
  - **LLM committee:** 4.420%
  - **MC benchmark:** 4.376%
- Dissents are essentially absent in both tracks.

In normal conditions, deliberation adds only a **modest hawkish wedge** of a few bps.

Political pressure: the center moves little, dissent rises a lot

- Mean outcome stays close to baseline:
  - **LLM committee:** 4.376%
  - **MC benchmark:** 4.386%
- But disagreement expands sharply:
  - **LLM dissent:** 88%
  - **MC dissent:** 61%

Political pressure shows up more in **consensus breakdown** than in large level shifts in the policy rate.

Payroll data revision: slightly more dovish, dissent present

- Softer labor data pushes both tracks lower:
  - **LLM committee:** 4.301%
  - **MC benchmark:** 4.315%
- Dissents remain elevated:
  - **LLM dissent:** 73.7%
  - **MC dissent:** 61.6%

New data matters, but much of its effect works through an already fragile consensus environment.

Validation

Unconditional Validation ("Chained")

- **Trajectory test (Feb 2000–Dec 2025):** 214 sequential meetings per chain; meeting t synthetic output becomes meeting $t+1$ input (errors compound).
- **Inputs & evaluation:** prior synthetic rate+statement, historical roster, lagged annual macro (FRED; Jan–Jun use prior-year)

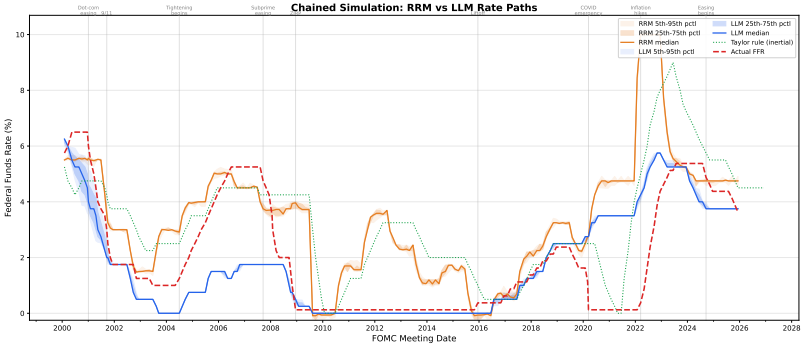


Figure: Synthetic paths vs actual (red) and Taylor (green).

Unconditional Validation (Chained): Results

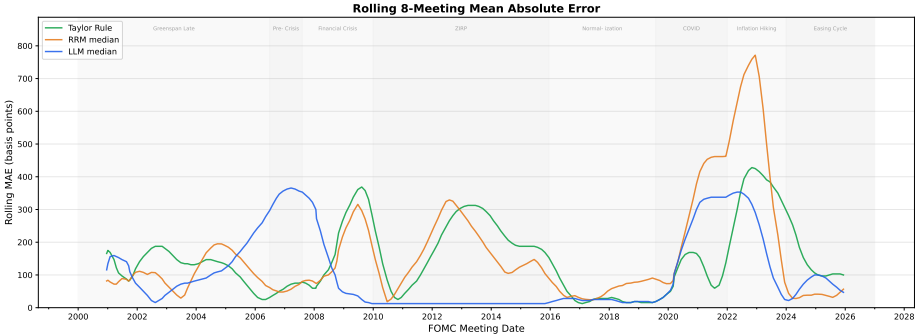


Figure: Rolling 8-meeting MAE: spikes at regime transitions; LLM is lowest for most of the sample.

We also run a one-step-ahead validation over 215 meetings (i.e., conditional), resetting each meeting to the true prior conditions yields highly accurate next meeting predictions: 9.1 bp MAE, 93% within 25 bp, and 77.2% direction accuracy.

Takeaways

- A synthetic FOMC lets us model **policy as a committee process**, not only as a reaction function.
- Under normal conditions, the LLM committee stays close to the benchmark.
- Under pressure, the main movement is in **dissent and consensus formation**, not just the rate level.
- This framework makes AI agents useful for **counterfactual policy analysis**:
  - chairs,
  - rosters,
  - agenda power,
  - political (and other) shocks,
  - data revisions,
  - etc.

Limitations

This work is best viewed as a structured simulation tool for committee behavior, not as a literal model of how policymakers think.

Some existing limitation:

- Priors and dissent costs are calibrated, not structurally estimated
- Personas rely on public information only
- Results, to some extent, remain sensitive to prompt design and model choice
- More tricky for historical data given look ahead bias in SOTA LLMs

Thank you!

Questions, comments, or suggestions?

Sophia Kazinnik
kazinnik@stanford.edu

Appendix

References I

Ashokkumar, A., Hewitt, L., Ghezae, I., and Willer, R. (2024). Predicting Results of Social Science Experiments Using Large Language Models.

Aubin Le Quéré, M., Schroeder, H., Randazzo, C., Gao, J., Epstein, Z., Perrault, S. T., Mimno, D., Barkhuus, L., and Li, H. (2024). Llms as research tools: Applications and evaluations in hci data work. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–7.

Bail, C. A. (2024). Can generative ai improve social science? *Proceedings of the National Academy of Sciences*, 121(21):e2314021121.

Dillion, D., Tandon, N., Gu, Y., and Gray, K. (2023). Can AI language models replace human participants? *Trends in Cognitive Sciences*, 27(7):597–600.

Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research.

Kosinski, M. (2023). Evaluating large language models in theory of mind tasks. *arXiv e-prints*, pages arXiv–2302.

References II

Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Willer, R., Liang, P., and Bernstein, M. S. (2024). Generative agent simulations of 1,000 people. *arXiv preprint arXiv:2411.10109*.