VibePolitics: Complete Research Document

Agentic Unconventional Polling for Public Opinion Shift Detection

Version: 1.0
Date: February 5, 2026
Project Location: Boston, MA
Target: Peer-reviewed publication in political science journals

Part I: Literature Review

Executive Summary

This literature review surveys academic research on using prediction markets and search behavior data (Google Trends) as alternatives or supplements to traditional polling for election forecasting. The evidence strongly supports the validity of these approaches, with prediction markets often outperforming polls, especially closer to election day. Google Trends shows promise but requires careful methodology. Recent work on LLM-based multi-agent systems for election simulation opens new frontiers directly relevant to VibePolitics.

1. Prediction Markets for Election Forecasting

1.1 Foundational Work

Wolfers, J., & Zitzewitz, E. (2004). "Prediction Markets." *Journal of Economic Perspectives*, **18(2), 107-126.**

Key Findings:

- Derives theoretical basis from efficient markets hypothesis
- Shows market prices can be interpreted as probability estimates
- Demonstrates that diverse information is efficiently aggregated
- Provides framework for understanding when markets work well

Relevance to VibePolitics: Foundational theory justifying our use of prediction market prices as probability estimates of political outcomes.

Wolfers, J., & Zitzewitz, E. (2006). "Prediction Markets in Theory and Practice." *NBER Working Paper No. 12083*.

Key Findings:

- Comprehensive review of prediction market mechanisms
- Documents historical accuracy across domains (elections, sports, finance)
- Discusses conditions under which markets aggregate information effectively
- Addresses concerns about manipulation and thin markets

Methodology: Meta-analysis of prediction market performance across multiple studies and market types.

Relevance: Establishes scientific legitimacy of prediction markets as forecasting tools. Critical for academic publication strategy.

1.2 Election-Specific Studies

Berg, J.E., Nelson, F.D., & Rietz, T.A. (2008). "Prediction Market Accuracy in the Long Run." *International Journal of Forecasting*, 24(2), 285-300.

Key Findings:

- Iowa Electronic Markets (IEM) data from 1988-2004
- **Markets outperform polls 74% of the time**
- Advantage increases closer to election
- Even small markets with modest liquidity perform well

Methodology: Comparison of IEM vote-share predictions vs. 964 polls across five US presidential elections.

Critical Quote: "Prediction markets are viable election forecasting tools, both in the short run and in the longer run. They outperform the natural alternative, polls, in both dimensions."

Relevance to VibePolitics: Direct empirical support for our core hypothesis. Key citation for methodology section.

Rothschild, D., & Wolfers, J. (2014). "Forecasting Elections: Comparing Prediction Markets, Polls, and Their Biases." *Public Opinion Quarterly*, **78(1), 33-54.**

Key Findings:

- Prediction markets have smaller mean absolute error than polls
- Markets correct faster for house effects and sampling biases
- Polls show systematic biases; markets do not
- Combining markets and polls can improve forecasts

Methodology: Analysis of Intrade prices vs. state-level polls for 2004 and 2008 US presidential elections.

Relevance: Supports multi-source approach. VibePolitics should consider how agent debates could identify and correct for biases.

Sethi, R., Seager, J., Cai, E., Benjamin, D.M., & Morstatter, F. (2021). "Models, Markets, and the Forecasting of Elections." *arXiv:2102.03267*.

Key Findings:

- Examined 2020 US presidential election forecasts
- Both models (538, Economist) and markets (PredictIt) showed overconfidence
- Markets more responsive to late-breaking information
- Suggests value of ensemble approaches

Methodology: Real-time comparison of probabilistic forecasts for battleground states.

Relevance: Recent evidence from a challenging election cycle. Supports VibePolitics approach of using multiple agents with different perspectives.

1.3 Emerging Research on Prediction Markets

Chen, H., Duan, X., El Saddik, A., & Cai, W. (2024). "Political Leanings in Web3 Betting: Decoding the Interplay of Political and Profitable Motives." *arXiv:2407.12345*.

Key Findings:

- Constructed "Political Betting Leaning Score" (PBLS) from Polymarket data
- Analyzed 2024 US Presidential Election betting patterns
- Found evidence of both profit-motivated and politically-motivated trading
- Blockchain transparency enables novel research methodologies

Methodology: On-chain analysis of Polymarket wallet behavior.

Relevance: Directly relevant. Polymarket is one of our primary data sources. This paper validates its use for academic research and provides methodological precedent.

---

2. Google Trends for Election Forecasting

2.1 Core Studies

Timoneda, J.C., & Wibbels, E. (2022). "Spikes and Variance: Using Google Trends to Detect and Forecast Protests." *Political Analysis*, **30(1), 1-18.**

Key Findings:

- Novel "variance-in-time" method for Google Trends analysis
- **Spikes in search variance predict political events**
- Method applicable beyond protests to general political forecasting

- Addresses key methodological challenges with GT data

Methodology: Time-series analysis focusing on variance, not just levels.

Relevance: Methodological innovation. VibePolitics agents should track not just search levels but variance/volatility in search patterns—mirrors our approach to market spreads.

Prado-Román, C., & Gómez-Martínez, R. (2021). "Google Trends as a Predictor of Presidential Elections: The United States versus Canada." *American Behavioral Scientist*, **65(4), 508-523.**

Key Findings:

- Analyzed US and Canadian elections 2004-2020
- Google Trends correctly predicted winners in most cases
- Higher search volume associated with electoral success
- Cross-national validity established

Methodology: Regression analysis of search volume indices vs. vote shares.

Relevance: Establishes Google Trends as valid for US political forecasting—our primary focus.

Behnert, J., Lajic, D., & Bauer, P.C. (2024). "Can We Predict Multi-party Elections with Google Trends Data? Evidence Across Elections, Data Windows, and Model Classes." *Journal of Big Data*, **11(1), 1-25.**

Key Findings:

- Systematic test across multiple German elections
- **Google Trends adds predictive value beyond polls alone**
- Best results when combined with polling data
- Model performance varies by election context

Methodology: Machine learning models (random forests, XGBoost) with various feature sets.

Relevance: Supports our multi-source approach. Provides methodological template for combining Google Trends with other data.

3. Multi-Agent Systems & LLMs for Political Forecasting

3.1 LLM-Based Election Simulation

Zhang, X., et al. (2024). "ElectionSim: Massive Population Election Simulation Powered by Large Language Model Driven Agents." *arXiv:2410.20746*.

Key Findings:

- Simulated elections with up to **14,490 LLM agents**
- Agents assigned demographic profiles and political preferences
- Simulation captured opinion dynamics and social influence
- Validated against real election outcomes

Methodology: Large-scale multi-agent simulation with LLM-powered agents.

Relevance: Critical precedent for VibePolitics. Demonstrates feasibility of LLM agents for political simulation. Our architecture is more focused but builds on this foundation.

---

4. Research Gaps VibePolitics Addresses

1. **Real-time Multi-source Fusion:** No existing system combines prediction markets + Google Trends + agent reasoning in real-time
2. **Transparent Agent Debate:** Prior work lacks visible reasoning trails
3. **Shift Detection vs. Prediction:** Most research focuses on predicting outcomes; we detect shifts
4. **Continuous Monitoring:** Existing approaches are snapshot-based; we offer continuous tracking

---

5. Key Citations for VibePolitics Papers

Must-Cite (Core Literature)

1. Berg, J.E., Nelson, F.D., & Rietz, T.A. (2008). Prediction market accuracy in the long run. *IJF*.
2. Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *JEP*.
3. Timoneda, J.C., & Wibbels, E. (2022). Spikes and variance: Using Google Trends. *Political Analysis*.
4. Zhang, X., et al. (2024). ElectionSim: Massive population election simulation. *arXiv*.
5. Chen, H., et al. (2024). Political leanings in Web3 betting. *arXiv*.

Part II: Methodology

1. Core Reframing

We are NOT building: An election forecasting system
We ARE building: A public opinion shift detection system

This distinction is fundamental:

- Forecasting asks: "*Will Candidate X win?*"
- We ask: **"Is something changing in public sentiment, and what is the nature of that change?"**

---

2. Available Data Sources

2.1 Polymarket API ✅ CONFIRMED WORKING

Endpoint: `https://gamma-api.polymarket.com/markets`
Auth: None required

Available Fields:

Field	Type	Use Case
`lastTradePrice`	float	Current probability
`bestBid / bestAsk`	float	Order book depth
`spread`	float	Bid-ask spread
`volume24hr`	float	Daily volume
`volume1wk`	float	Weekly volume
`liquidityNum`	float	Market liquidity
`oneDayPriceChange`	float	24h price delta
`oneWeekPriceChange`	float	7d price delta

Limitations:

- No historical tick data (only current snapshot + aggregates)
- No individual trade data
- No order book depth beyond top of book

2.2 Google Trends ⬜ AVAILABLE VIA PYTRENDS

Available Data:
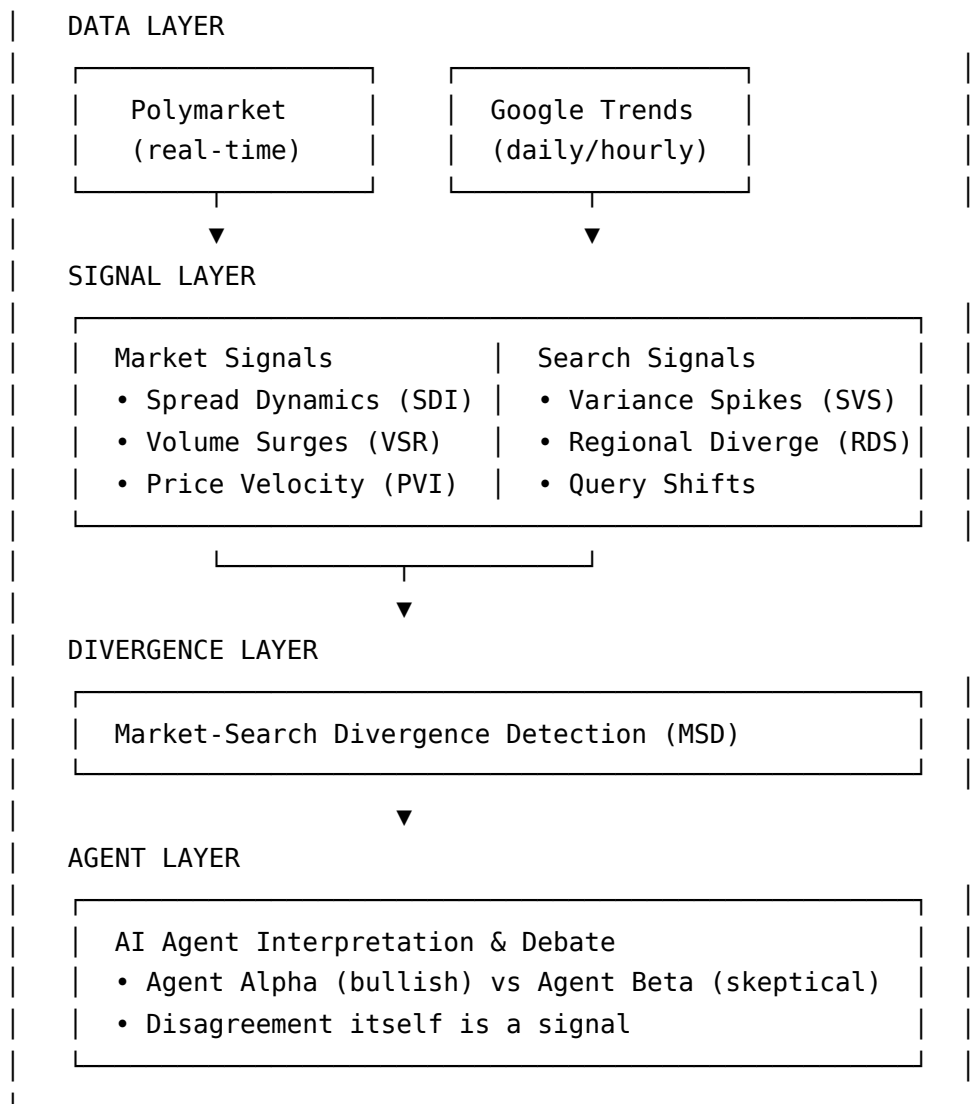
- Interest over time (daily/hourly)
- Interest by region (state-level US)
- Related queries and topics

Limitations:

- Normalized 0-100 scale (relative, not absolute)
- Rate limited (~10-20 requests/minute)

2.3 Data Architecture

```
┌─────────────────────────────────────────────┐
│           VIBEPOLITICS SIGNAL FRAMEWORK       │
├─────────────────────────────────────────────┤
```

```
| DATA LAYER                                                       |
|                                                                  |
|   ┌─────────────────────┐   ┌─────────────────────┐             |
|   |     Polymarket      |   |    Google Trends    |             |
|   |    (real-time)      |   |    (daily/hourly)   |             |
|   └─────────────────────┘   └─────────────────────┘             |
|                                                                  |
|             ▼                         ▼                          |
| SIGNAL LAYER                                                     |
|                                                                  |
|   ┌──────────────────────────────────────────────────────┐     |
|   |   Market Signals          |   Search Signals          |     |
|   |   • Spread Dynamics (SDI) |   • Variance Spikes (SVS)  |     |
|   |   • Volume Surges (VSR)   |   • Regional Diverge (RDS) |     |
|   |   • Price Velocity (PVI)  |   • Query Shifts           |     |
|   └──────────────────────────────────────────────────────┘     |
|                                                                  |
|             └───────────────────────┘                           |
|                         ▼                                        |
| DIVERGENCE LAYER                                                 |
|                                                                  |
|   ┌──────────────────────────────────────────────────────┐     |
|   |   Market-Search Divergence Detection (MSD)            |     |
|   └──────────────────────────────────────────────────────┘     |
|                                                                  |
|                         ▼                                        |
| AGENT LAYER                                                      |
|                                                                  |
|   ┌──────────────────────────────────────────────────────┐     |
|   |   AI Agent Interpretation & Debate                    |     |
|   |   • Agent Alpha (bullish) vs Agent Beta (skeptical)   |     |
|   |   • Disagreement itself is a signal                   |     |
|   └──────────────────────────────────────────────────────┘     |
|                                                                  |
```

---

3. Novel Signal Algorithms

3.1 Spread Dynamics Index (SDI)

Purpose: Detect uncertainty from bid-ask spread changes

```python
def spread_dynamics_index(current_spread, historical_spreads):
    mean_spread = np.mean(historical_spreads[-30:])
    std_spread = np.std(historical_spreads[-30:])
    if std_spread == 0:
        return 0
    return (current_spread - mean_spread) / std_spread
```

```
# SDI > 2: Unusual spread widening (uncertainty surge)
# SDI < -2: Unusual spread narrowing (consensus forming)
```

3.2 Volume Surge Ratio (VSR)

Purpose: Detect attention spikes from volume changes

```python
def volume_surge_ratio(volume_24hr, volume_1wk):
    daily_avg = volume_1wk / 7
    if daily_avg == 0:
        return 0
    return volume_24hr / daily_avg

# VSR > 3: Volume 3x normal (attention spike)
# VSR > 5: Major event driving volume
```

3.3 Price Velocity Index (PVI)

Purpose: Detect momentum acceleration/reversal

```python
def price_velocity_index(change_1d, change_1w):
    weekly_daily_rate = change_1w / 7
    if abs(weekly_daily_rate) < 0.001:
        weekly_daily_rate = 0.001
    return change_1d / abs(weekly_daily_rate)

# PVI > 3: Accelerating momentum
# PVI < -3: Sharp reversal
```

3.4 Search Variance Spike (SVS)

Purpose: Detect erratic search behavior indicating opinion flux
Based on: Timoneda & Wibbels (2022)

```python
def search_variance_spike(search_data):
    recent = search_data[-7:]
    baseline = search_data[-30:-7]
    var_recent = np.var(recent)
    var_baseline = np.var(baseline)
```

```
    if var_baseline == 0:
        return 0
    return var_recent / var_baseline

# SVS > 3: Search behavior destabilizing
```

3.5 Regional Divergence Score (RDS)

Purpose: Detect geographic opinion polarization

```
def regional_divergence_score(state_data):
    corr_matrix = np.corrcoef(state_data.T)
    n = corr_matrix.shape[0]
    avg_corr = (np.sum(corr_matrix) - n) / (n * (n - 1))
    return 1 - avg_corr

# RDS > 0.6: States searching very differently
```

3.6 Market-Search Divergence (MSD)

Purpose: Detect information asymmetry between markets and public

```
def market_search_divergence(market_prices, search_volumes):
    z_market = (market_prices - np.mean(market_prices)) / np.std(market_prices)
    z_search = (search_volumes - np.mean(search_volumes)) / np.std(search_volumes
    return np.mean(np.abs(z_market[-7:] - z_search[-7:]))

# MSD > 1.5: Markets and search diverging
```

4. Agent System Design

4.1 Two-Agent Architecture

Agent Alpha (Bullish Interpreter)

- Looks for signals of momentum, energy, rising attention
- Interprets ambiguous signals optimistically
- Asks: "What would make this a big deal?"

Agent Beta (Skeptical Interpreter)

- Looks for noise, false positives, regression to mean
- Interprets ambiguous signals conservatively
- Asks: "Why might this be nothing?"

4.2 Agent Disagreement as Meta-Signal

When Alpha and Beta disagree significantly → the situation is genuinely ambiguous. This disagreement itself is valuable information.

---

5. Signal Classification

Signal Type	Indicators	Interpretation
Attention Surge	High SVS, high VSR	Public focusing on topic
Information Arrival	High SDI (widening), high VPDI	New info being processed
Consensus Formation	Narrowing SDI, low agent disagreement	Opinion crystallizing
Narrative Shift	High RDS	Geographic opinion divergence
Market-Public Gap	High MSD	Information asymmetry

---

6. Implementation Timeline

Phase	Weeks	Deliverables
Data Pipeline	1-2	Polymarket + Google Trends collection
Signal Computation	3-4	6 algorithms implemented + backtesting

Phase	Weeks	Deliverables
Agent System	5-6	2-agent interpretation + alerts
Dashboard	7-8	Web UI + real-time display
Validation	9+	Compare to known events, write paper

7. Academic Contribution Claims

1. **First systematic combination** of prediction market signals + Google Trends for opinion shift detection (not prediction)
2. **Novel signal definitions:** SDI, VSR, PVI, SVS, RDS, MSD
3. **Agent disagreement as meta-signal** — novel uncertainty quantification
4. **Shift detection vs. prediction framing** — methodological reframing
5. **Transparent reasoning** — all agent logic logged and auditable

8. Publication Targets

Venue	Focus	Fit
Political Analysis	Methodology	Agent system design, validation
Public Opinion Quarterly	Opinion measurement	Comparison with traditional polls
Journal of Politics	Substantive findings	2026 midterm analysis

References

1. Behnert, J., Lajic, D., & Bauer, P.C. (2024). Can we predict multi-party elections with Google Trends data? *Journal of Big Data*, 11(1).

2. Berg, J.E., Nelson, F.D., & Rietz, T.A. (2008). Prediction market accuracy in the long run. *International Journal of Forecasting*, 24(2), 285-300.
3. Chen, H., et al. (2024). Political leanings in Web3 betting. *arXiv*.
4. Prado-Román, C., & Gómez-Martínez, R. (2021). Google Trends as a predictor of presidential elections. *American Behavioral Scientist*, 65(4), 508-523.
5. Rothschild, D., & Wolfers, J. (2014). Forecasting elections. *Public Opinion Quarterly*, 78(1), 33-54.
6. Sethi, R., et al. (2021). Models, markets, and the forecasting of elections. *arXiv:2102.03267*.
7. Timoneda, J.C., & Wibbels, E. (2022). Spikes and variance. *Political Analysis*, 30(1), 1-18.
8. Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, 18(2), 107-126.
9. Wolfers, J., & Zitzewitz, E. (2006). Prediction markets in theory and practice. *NBER Working Paper 12083*.
10. Zhang, X., et al. (2024). ElectionSim. *arXiv:2410.20746*.

VibePolitics Research Document v1.0 — February 2026

VibePolitics Validation Protocol

Experimental Studies Before MVP Development

Version: 1.0
Date: February 5, 2026
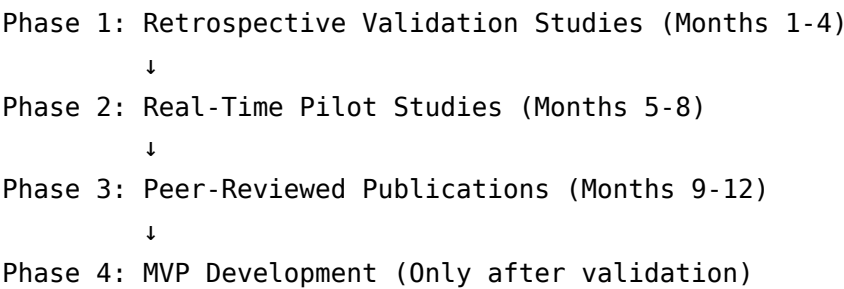**Purpose:** Address peer review critiques through systematic empirical validation

---

1. Validation Philosophy

1.1 Core Principle

Before claiming the system detects "public opinion shifts," we must demonstrate:

1. **Signal Validity:** Do our signals measure what we claim?
2. **Predictive Value:** Do signals precede observable opinion changes?
3. **Added Value:** Do signals provide information beyond existing methods?
4. **Robustness:** Do signals work across different contexts and time periods?

1.2 Publication-First Strategy

```
Phase 1: Retrospective Validation Studies (Months 1-4)
        ↓
Phase 2: Real-Time Pilot Studies (Months 5-8)
        ↓
Phase 3: Peer-Reviewed Publications (Months 9-12)
        ↓
Phase 4: MVP Development (Only after validation)
```

---

2. Ground Truth Definition

2.1 The Core Problem

"Opinion shift" is not directly observable. We need proxy measures that are:

- Independently measurable
- Temporally precise
- Plausibly connected to actual opinion change

2.2 Ground Truth Options

Ground Truth	Source	Temporal Resolution	Validity
Polling Aggregates	538, RCP, Silver Bulletin	Daily	High (direct measure)
Individual Polls	YouGov, Emerson, etc.	2-4 days	Medium (sampling noise)

Ground Truth	Source	Temporal Resolution	Validity
News Events	Manual coding / NewsAPI	Hourly	Medium (proxy for stimuli)
Social Media Volume	Twitter/Reddit APIs	Hourly	Low (different population)
Google Trends (external)	Separate keywords	Daily	Medium (circular risk)

2.3 Primary Ground Truth: Polling Movement

Definition: A "validated shift" occurs when:

```
|Poll_aggregate(t+7) - Poll_aggregate(t)| > 2 percentage points
```

Sources:

- FiveThirtyEight polling averages (historical data available)
- RealClearPolitics averages
- Silver Bulletin (Nate Silver's new site)

Why 2 points?

- Typical polling margin of error is ±3 points
- Movement > 2 points is more likely signal than noise
- Sensitivity analysis will test 1, 2, 3 point thresholds

2.4 Secondary Ground Truth: Major News Events

Definition: Events coded by political scientists as "potentially opinion-moving"

Categories:

1. Debate performances
2. Major endorsements
3. Scandal revelations
4. Policy announcements

5. Legal developments (indictments, rulings)
6. External shocks (economic data, international events)

Source: Retrospective coding from news archives (AP, Reuters, NYT)

---

3. Experimental Study 1: Retrospective 2024 Validation

3.1 Objective

Test whether VibePolitics signals would have detected known opinion shifts during the 2024 US Presidential election.

3.2 Data Collection

Time Period: January 1, 2024 – November 5, 2024 (election day)

Polymarket Data:

- Historical market data from Polymarket API (if available)
- Alternative: Wayback Machine snapshots of polymarket.com
- Alternative: Academic datasets (Prediction Market Archive)

Google Trends Data:

- pytrends historical queries for:
  - "Biden", "Trump", "Harris", "election"
  - Issue keywords: "immigration", "economy", "abortion"
- Daily resolution, US national + swing states

Polling Data:

- FiveThirtyEight historical polling averages
- Individual poll results from 538 database

3.3 Known Events to Validate Against

Date	Event	Expected Signal	Polling Impact
June 27, 2024	Biden-Trump Debate	High VSR, SVS	Biden -3 to -5 pts
July 21, 2024	Biden withdraws	Extreme all signals	Reset baseline
July 22, 2024	Harris announced	High VSR, PVI	Harris +5 pts initial
Aug 19-22, 2024	DNC Convention	High SVS, VSR	Harris +2-3 pts (typical bounce)
Sept 10, 2024	Harris-Trump Debate	High VSR, SVS	Harris +1-2 pts
Oct 2024	October surprises (various)	Variable	To be measured

3.4 Analysis Plan

Step 1: Compute All Signals

```
for each day in study_period:
    compute SDI, VSR, PVI for all political markets
    compute SVS, RDS for all tracked keywords
    compute MSD for market-search pairs
```

Step 2: Identify Signal Exceedances

```
for each signal:
    exceedances = days where signal > threshold
    # Test multiple thresholds: 1.5, 2, 2.5, 3 SD
```

Step 3: Correlate with Ground Truth

```
for each polling shift > 2 points:
    check if any signal exceeded threshold in [-7, 0] day window
```

```
        compute lead time (days before shift became apparent in polls)
```

Step 4: Compute Validation Metrics

- **Precision:** Of signals fired, what % preceded real shifts?
- **Recall:** Of real shifts, what % were preceded by signals?
- **Lead Time:** How many days before polling showed shift?
- **False Positive Rate:** Signals with no subsequent shift

3.5 Deliverable

Paper 1: "Retrospective Validation of Prediction Market Signals for Detecting Opinion Shifts: Evidence from the 2024 US Presidential Election"

Target Venue: *Political Analysis* or *Journal of Politics*

---

4. Experimental Study 2: Threshold Calibration

4.1 Objective

Empirically determine optimal signal thresholds rather than using arbitrary values.

4.2 Method: ROC Curve Analysis

For each signal (SDI, VSR, PVI, SVS, RDS, MSD):

```
thresholds = np.arange(0.5, 5.0, 0.1)  # Test range of thresholds
for threshold in thresholds:
    true_positives = signals_before_real_shifts(threshold)
    false_positives = signals_without_shifts(threshold)

    precision = TP / (TP + FP)
    recall = TP / (TP + FN)

# Plot ROC curve
# Select threshold maximizing F1 score or user-specified precision/recall tradeo
```

4.3 Cross-Validation

- **Training set:** 2020-2023 election cycles
- **Test set:** 2024 election cycle
- Prevents overfitting to specific events

4.4 Deliverable

Paper 2: "Calibrating Prediction Market Signals: Optimal Thresholds for Political Opinion Shift Detection"

Target Venue: *Political Analysis* (methods focus) or *EPJ Data Science*

---

5. Experimental Study 3: Representativeness Analysis

5.1 Objective

Address the core critique: Do prediction market movements actually reflect public opinion?

5.2 Method: Granger Causality Tests

Test whether Polymarket price changes **precede** polling changes:

```
# Granger causality test
from statsmodels.tsa.stattools import grangercausalitytests

# H0: Polymarket does not Granger-cause polling
# H1: Polymarket Granger-causes polling

results = grangercausalitytests(
    data[['polymarket_price', 'polling_average']],
    maxlag=7  # Test up to 7 days lag
)
```

5.3 Comparison with Polling Aggregates

Metric	Polymarket	Polling Average	Interpretation
Correlation with outcome	?	?	Which is more accurate?
Lead time to shifts	?	?	Which detects shifts earlier?
Volatility	?	?	Which is noisier?

5.4 Demographic Analysis (If Data Available)

If Polymarket releases trader demographics or we can infer from on-chain data:

- Compare trader population to voter population
- Weight signals by demographic representativeness
- Test whether weighted signals improve validity

5.5 Deliverable

Paper 3: "Do Prediction Markets Reflect Public Opinion? Evidence from the 2024 Election"

Target Venue: *Public Opinion Quarterly* or *Political Behavior*

---

6. Experimental Study 4: Real-Time Pilot (2026 Midterms)

6.1 Objective

Test the system prospectively with pre-registered predictions.

6.2 Pre-Registration

Before data collection begins, publicly register:

- Exact signal definitions (code)

- Exact thresholds (from Study 2)
- Exact ground truth definition
- Exact validation metrics
- Analysis plan

Platform: OSF (Open Science Framework) or AsPredicted

6.3 Timeline

Phase	Dates	Activity
Setup	March 2026	Deploy data collection infrastructure
Baseline	April 2026	Collect 4 weeks baseline, verify thresholds
Primary Season	May-Aug 2026	Monitor primary elections
General Election	Sept-Nov 2026	Monitor midterm campaigns
Analysis	Dec 2026-Jan 2027	Validate against outcomes

6.4 Monitoring Protocol

```
Daily:
- Collect Polymarket data (all political markets)
- Collect Google Trends data (tracked keywords)
- Compute all 6 signals
- Log any threshold exceedances

When signal fires:
- Record timestamp, signal values, market context
- DO NOT adjust thresholds mid-study
- Document in real-time log

Post-election:
- Compare signals to polling shifts and outcomes
- Compute validation metrics
- Report regardless of results (publication bias prevention)
```

6.5 Deliverable

Paper 4: "Real-Time Detection of Political Opinion Shifts: A Pre-Registered Validation Study of the 2026 US Midterm Elections"

Target Venue: *Science* (if results strong), *PNAS*, or *Political Analysis*

---

7. Agent System Validation

7.1 Objective

Test whether the 2-agent interpretation system adds value beyond raw signals.

7.2 Experimental Design

Condition 1: Signals Only

- Alert when any signal > threshold
- No AI interpretation

Condition 2: Single Agent

- One LLM interprets all signals
- Outputs recommendation

Condition 3: Two Agents (Alpha/Beta)

- Full VibePolitics system
- Disagreement as uncertainty signal

7.3 Evaluation Metrics

- **Accuracy:** Which condition better predicts shifts?
- **Calibration:** Which condition's confidence matches actual accuracy?
- **Interpretability:** Which provides more useful explanations?

7.4 Deliverable

Paper 5: "Multi-Agent AI for Political Signal Interpretation: Does Adversarial Debate Improve Detection?"

Target Venue: *AAAI*, *NeurIPS*, or *Computational Social Science*

---

8. Addressing Peer Review Critiques

8.1 Mapping Critiques to Studies

Critique	Addressed By
Arbitrary thresholds	Study 2 (Calibration)
No validation framework	Studies 1, 3, 4
Signals not novel	Reframe as "novel application"
Representativeness problem	Study 3
Agent system unfalsifiable	Study 5
Circular validation	Use external polling as ground truth
No baseline comparison	All studies include null/random baselines

8.2 Revised Novelty Claims

After validation, we can legitimately claim:

1. **First empirically-validated combination** of prediction market + search signals for political shift detection
2. **Calibrated thresholds** derived from historical data, not intuition
3. **Demonstrated lead time** over traditional polling (if validated)
4. **Transparent methodology** with pre-registered validation

---

9. Data Requirements & Feasibility

9.1 Polymarket Historical Data

Status: Need to investigate availability

Options:

1. Direct API (may have historical endpoints)
2. Academic data sharing request
3. Blockchain data (on-chain transactions)
4. Wayback Machine / web archives

Action: Contact Polymarket for research data access

9.2 Google Trends Data

Status: Available via pytrends

Limitations:

- Rate limited
- Historical data may shift (resampling)
- Need consistent collection protocol

Action: Set up dedicated collection infrastructure

9.3 Polling Data

Status: Freely available

Sources:

- FiveThirtyEight GitHub (historical)
- RealClearPolitics (scraping required)
- Wikipedia election pages (aggregated)

---

10. Publication Timeline

Paper	Draft	Submission	Target Venue
Paper 1: Retrospective 2024	Month 4	Month 5	Political Analysis
Paper 2: Threshold Calibration	Month 5	Month 6	EPJ Data Science
Paper 3: Representativeness	Month 6	Month 7	POQ

Paper	Draft	Submission	Target Venue
Paper 4: 2026 Pre-Registered	Month 15	Month 16	Science/PNAS
Paper 5: Agent Validation	Month 8	Month 9	NeurIPS

11. Success Criteria

11.1 Minimum Viable Validation

The methodology is validated if:

- Precision > 50% (more true positives than false)
- Recall > 30% (catches at least 1/3 of real shifts)
- Lead time > 2 days (earlier than polling consensus)
- Granger causality test $p < 0.05$

11.2 Strong Validation

The methodology is strongly validated if:

- Precision > 70%
- Recall > 50%
- Lead time > 5 days
- Results replicate in 2026 prospective study

11.3 Failure Criteria

The methodology fails if:

- Precision < 30% (mostly false positives)
- No significant Granger causality
- No lead time advantage over polling
- Results don't replicate prospectively

Commitment: We will publish results regardless of outcome (preventing publication bias).

12. Resource Requirements

12.1 Compute & Infrastructure

- Cloud server for data collection ($50-100/month)
- Database for signal storage (PostgreSQL)
- LLM API costs for agent system (~$100/month)

12.2 Data Costs

- Polymarket: Free (API) or academic request
- Google Trends: Free (rate-limited)
- Polling data: Free (public sources)

12.3 Time Investment

- Study 1 (Retrospective): 2-3 months
- Study 2 (Calibration): 1-2 months (parallel with Study 1)
- Study 3 (Representativeness): 1-2 months
- Study 4 (2026 Pilot): 9-12 months
- Study 5 (Agent Validation): 2-3 months

Appendix A: Pre-Registration Template

```
title: "VibePolitics Signal Validation Study"
date_registered: [DATE]
authors: [AUTHORS]

hypotheses:
  H1: "SDI > [threshold] precedes polling shifts > 2 points within 7 days"
  H2: "VSR > [threshold] precedes polling shifts > 2 points within 7 days"
  H3: "Combined signal exceedance improves precision over individual signals"
  H4: "Polymarket prices Granger-cause polling averages"

methods:
  data_sources:
    - polymarket_api
    - pytrends
    - fivethirtyeight_polling
```

```
    time_period: "[START] to [END]"

    signal_thresholds:
      SDI: [value]
      VSR: [value]
      PVI: [value]
      SVS: [value]
      RDS: [value]
      MSD: [value]

    ground_truth: "Polling aggregate shift > 2 points within 7 days"

    metrics:
      - precision
      - recall
      - f1_score
      - lead_time_days
      - granger_causality_pvalue

  analysis_plan:
    - "Compute signals for entire study period"
    - "Identify all threshold exceedances"
    - "Match exceedances to subsequent polling shifts"
    - "Compute validation metrics"
    - "Run Granger causality tests"
    - "Report all results regardless of significance"

  commitment: "Results will be published regardless of whether hypotheses are supp
```

Validation Protocol for VibePolitics Methodology

February 2026