

What the news says vs. what Polymarket prices: 257 markets, nine months, 751k articles

Thinking Text LLC, 2026-10-03

Reader: spinf-12b (Gemma 4 12B base) via the spinf scoring API · Queries: Claude Opus 5.5

Abstract

We read 751,370 English news articles from 2026-01-01 to 2026-10-02 against 257 questions covering the 200 most traded open Polymarket events. 238 of those questions were drafted by an LLM and 19 built by hand. Each market becomes one question with mutually exclusive outcomes priced from its order book. A keyword filter routes articles to it, and a 12B base language model reads each article three ways, scoring every outcome against the same prompt with no article.

In many markets, the news and the price move together, sometimes strikingly closely. On the 84 markets with twelve or more weeks of solid data, a 14-day average of the news read correlates with the price above 0.5 in 27% of markets and above 0.7 in 14%. Pairing each market's news with another market's price over the same dates gives 1% and 0%. The best reach 0.85–0.90 (Georgia Senate, the town of Lyman, a US–Iran deal clause, Anthropic vs OpenAI to IPO first). The relationship is also strikingly uneven: 30% of markets correlate negatively. The price and the news move within days of each other, with neither systematically ahead.

Rankings carry information in layers. With no article at all, the model already puts the market's favourite first on 54% of market-days (chance 41%). Reading the articles raises that to 59%, and the articles change the model's favourite on 21% of days. Some markets agree almost perfectly (Hormuz shipping, a Russia–Ukraine ceasefire: 90–94% of outcome pairs in the market's order). Others disagree just as sharply (the October Fed meeting: 6%).

Pooled over every market, the news also carries a small forward signal: a partial correlation of +0.073 with the next day's price change beyond the market's own reversion (permutation null +0.031, clustered $t = 9.7$). It grew through the year. We did not test a trading strategy.

The questions were drafted after the period they are tested on (Section 7), so the live feed, running since 2026-10-03, is the out-of-sample test. The computed data and that feed, updated hourly, are freely accessible with a free API key at thinkingtext.com/polymarket-news-index. We do not publish the per-market queries.

What we found, briefly

- **News and prices track each other in a large minority of markets.** Figure 3 is the distribution. Figure 4 shows the ten closest markets so you can judge the fit yourself.
- **The variation between markets is the most interesting**

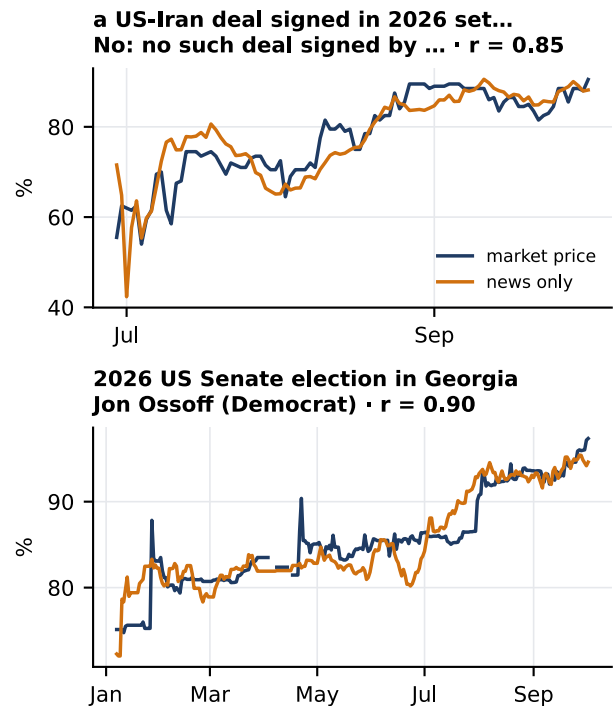


Figure 1: Two of the closest markets: a clause of a possible US–Iran deal, and the Georgia Senate race. Blue: the market price. Orange: the news alone, read from the day's articles with no market data (EMA, half-life 14 days), rescaled to the price's range; r is the correlation over the market's history. Section 4.1 has the full distribution and the top ten.

result. It is not explained by one smoothing window, by article volume, or by category alone. Why does Georgia's Senate race track at 0.90 while another race doesn't? We don't know yet, and it is the obvious next question.

- **A 12B model's prior is a decent forecaster of the favourite,** before reading anything, and reading the news moves it toward the market. Where it disagrees strongly (Fed, Bitcoin), the disagreement is mostly the model's prior, not the articles. Whether that is a reader artefact or a real divergence is open.
- **Asking the model “will X go up?” does not work.** Asking which outcome the article supports, relative to an empty article, does.
- **Narrow keyword filters beat broad ones:** 14 articles a day per market forecast better than 77 or 174.
- **An LLM writes good queries:** Claude drafts a market in about a minute, matching or beating our hand-built versions. Letting it revise from six months of its own

results did not help further.

1. Goal

Polymarket prices are a consensus. Traders who use them already have a view; what they lack is a cheap way to see what the news actually says about each outcome, whether that matches the price, and what changed today. We built a measurement layer for that:

1. per market and outcome, what the day's articles point to, against what the market prices;
2. how that changes over hours, days and weeks;
3. the articles behind each change, as links.

We expected the news and the market to agree, since most news covers facts the market already prices. What we wanted to learn is where they agree, where they don't, and whether the differences mean anything.

2. Data

Markets. Both Polymarket exchanges: polymarket.com (public gamma and data APIs: 11,383 events, 126,882 markets, 1.57M daily price bars since 2025) and Polymarket US (public gateway and Daily Market Reports: 7,747 events, 95,985 contracts). Sports and automated markets are excluded. Every day we rank the open events of both exchanges by 30-day traded dollars, counted from the trade tape rather than the API's monthly-volume field, which stays high for markets that stopped trading months ago. The top 200 are read. A daily close is the price at 00:00 UTC.

News. GDELT's Web News NGrams 3.0 (every 15 minutes, ~5 GB of n-grams a day, with their URLs), reconstructed into article text for English articles from 2,654 outlets:

- 108 national and wire outlets;
- 189 trade and finance outlets;
- 2,357 outlets selected for this project from a census of every English domain in the 2025 data. A domain is kept if it publishes at least two relevant articles a day and at least 15% of its articles are relevant (they name a market's subject, or two terms of one topic), with under 30% sports and under 80% duplicated copy.

That is ~70,000 English articles a day. Every article is cut to its first 260 words (title and lead), deduplicated by canonical URL, and clustered by MinHash: one read per cluster, with the number of outlets that carried it. Aggregators are dropped, and so are three templated "content farm" domains whose copy is ~60% auto-generated ticker text.

3. Method

3.1 One market, one question

A Polymarket "event" is a set of yes/no contracts. Each becomes one question whose outcomes are mutually exclusive and exhaustive, each priced by a formula over the contracts:

- **contenders** (who wins): one outcome per candidate priced at 2% or more, plus "someone else" = $1 - \Sigma$;
- **by-date** (when): windows priced by differencing the cumulative contracts ("by Oct 31", "by Nov 30" → "in Novem-

ber" = $P(\text{by Nov 30}) - P(\text{by Oct 31})$);

- **ranges** (how many / how high): bins from differenced threshold contracts;
- **binary**: yes / no.

Prices are renormalised over the outcomes still open each day. Appendix B shows five markets exactly as configured.

3.2 Which articles a market reads

Each market has a keyword filter: an OR of rules, a rule an AND of term groups, a group an OR of terms (case-folded, accents folded, whole words; =Fed is case-sensitive). A rule can be required to match within one sentence. The filter runs on the same 260 words the model reads. A median market reads 10 articles a day; the busiest, several hundred.

3.3 Three reads per article, against an empty article

The reader is a base model behind a scoring API: send a prompt and a list of continuations, get the log-probability of each. Per article and market, three prompts share the article prefix (the API reuses it, so the extra prompts are cheap):

- **mc**: the outcomes as lettered options; "the most likely answer is option" → $\log P(A), \log P(B), \dots$
- **yn**: per outcome, "is the answer "{outcome}"? Answer:" → $\log P(\text{Yes}) - \log P(\text{No})$
- **dir**: per outcome, "the chance that the answer is "{outcome}" is likely to" → $\log P(\text{increase}) - \log P(\text{decrease})$

Every prompt starts with "Today is {date}." Each read is measured against the same prompt with the article replaced by "N/A", on the same date. That removes the model's prior for each label and letter, and makes the read deadline-aware: the prior for "in October" on October 25 is not its prior on June 1. Per market we subtract the mean over its outcomes, so a read means "this outcome, against this market's other outcomes". The three reads are standardised and averaged (**comb**). This is the **news read** used everywhere below unless stated. Templates: Appendix A.

3.4 News-implied odds and flags

For the feed, each outcome gets news-implied odds. Take the market's own 30-day mean relative log-odds for the outcome, shift it by $b \times$ (the outcome's 7-day news average minus its 30-day mean), and softmax over the open outcomes. b is one pooled scale fitted on past data only (0.27 in October). The odds answer "where would the price sit if it followed the news". Because they are anchored on the market, **no chart or correlation in Section 4.1 uses them**: those compare the news read alone with the price. The flag compares implied odds with the price: *news more positive* at $1.25\times$ or more, *news less positive* at $0.8\times$ or less.

3.5 Writing the queries with an LLM

Writing filters and outcomes by hand does not scale to hundreds of markets that change weekly. Claude Opus 5.5 gets:

- our guidelines ($\approx 2,800$ words: structures, filter patterns, past mistakes);
- the market's data (description, contracts, resolution texts, prices).

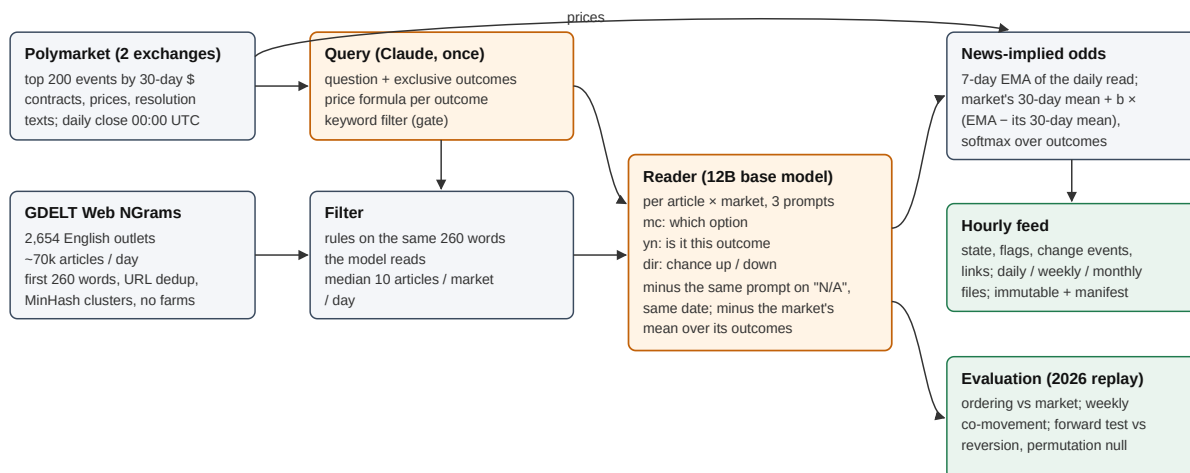


Figure 2: The pipeline. Claude writes each market’s question, outcomes and filter once; every day the filtered articles are read three ways against an empty article; the reads become the news series, news-implied odds, the hourly feed, and the evaluation.

It returns the question, the outcomes with price formulas, and the filter. Validation is automatic:

1. the contract slugs exist;
2. the outcome prices are non-negative and sum to ~ 1 ;
3. the filter runs on the last 7 days of articles (volume);
4. a cheap second Claude call judges whether 30 matched excerpts bear on the outcome.

On a failure, Claude gets the report and revises once. A market takes about a minute and 25–110k input tokens. Claude declined 2 of the 200 markets: one turns on a leaderboard score no article reports, the other on near-zero-priced exchange listings.

4. Results

4.1 Where news and price move together

For each market and outcome we smooth the daily news read causally and correlate it with the market’s relative log-odds over the market’s history. We used 7 smoothings: an EMA with half-life 3, 7, 14 or 30 days, or a trailing average over 7, 14 or 30 article days. Each is used either as a level or as a trailing z-score over 30, 60 or 90 days, as in our oil index, which gives 28 settings in total. A market’s correlation is the mean over its outcomes. We require 15+ article days and prices on 60+ days and on 80% of the market’s days, since thin markets trade in bursts.

The null matters. Two smooth, trending series correlate by accident. For every market we also correlate its news with the price of an outcome of a *different* market over the same dates (three random pairings each). That shows how often a high correlation appears from trends alone.

(101 markets with 60+ days; Figure 3 restricts to the 84 with 12+ weeks.) The level settings find the most markets above 0.5. The 60–90-day z-scores have the highest median, because they remove slow drifts. Short z-score windows (30 days) lose most of the relationship. Whatever the setting, the null stays near zero. The result does not hinge on one tuned window.

What stands out:

- **Elections and company races track tightly:** Georgia

Senate 0.90, California governor 0.79, House control 0.73, Anthropic vs OpenAI to IPO first 0.83.

- **Some geopolitical markets do too:** the town of Lyman (0.85), the Iranian regime (0.72), whether Reza Pahlavi enters Iran (0.81). These include big trend reversals, such as the late-September collapse of the Lyman market, which the news follows.
- **Even a box-office market qualifies:** the year’s biggest opening weekend, 0.68.

Who leads? Shifting the price against the news by up to four weeks either way, the median correlation peaks at the same day (0.218), and is 0.204 a week earlier and 0.199 a week later. On the 23 markets above 0.5 it peaks within three days (0.727 three days earlier, 0.719 the same day). News and price move together within days. Daily data cannot tell which reacts first, and the timing tests in Section 6 found no faster read.

The other side of the distribution is just as interesting. 30% of markets correlate *negatively*, some at -0.3 to -0.7 . Possible causes include filters that catch counter-narratives, questions the model reads backwards, markets that move on information that doesn’t reach English news, and markets that react against the news. We have not separated these, and each would teach us something.

4.2 Ranking: what the model believes, what the articles add, what the market says

Per market-day with 3+ articles, we compare the order of the outcomes in three readings: the model with no article (only the question and the date), the model reading the articles (the mean of its multiple-choice answers), and the market.

(18,497 market-days.) The model’s prior is already a better-than-chance forecaster of the favourite. It knows that cease-fires rarely come “this year” and that incumbents usually hold. The articles change its favourite on 21% of market-days and move it toward the market.

Agreement grows with the market’s confidence. With the favourite priced above 80%, the news picks the same favourite on 77% of days (chance 47%). With no outcome above 50%, it does so on 30% (chance 18%).

period	2026-01-01 – 2026-10-02 (275 days)
questions read	257 (221 with at least one day of 3+ articles and prices)
articles read (unique)	751,370
(article, market) pairs	1,691,926
(article, market, prompt) queries	5,075,778
billed tokens, all runs incl. experiments	≈ 2.5 billion
throughput	~43k billed tokens/s on two GPUs

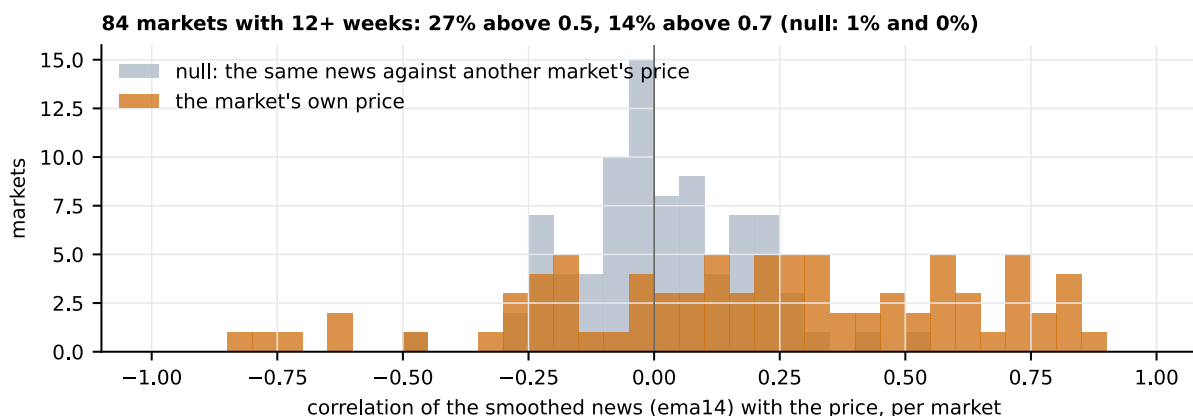


Figure 3: Per market (84 markets with 12+ weeks), the correlation of the news read (EMA, half-life 14 days) with the price, against the same news paired with another market’s price. 27% of markets are above 0.5 and 14% above 0.7; the null gives 1% and 0%.

Where the rankings disagree, the disagreement is mostly the model’s prior. On the October Fed meeting, the articles’ answer favours “cut” (60%) while the market prices “hold”. With no article, the model gives “cut” 64%. On Bitcoin back above \$100k, the articles favour “before October 1” (33%) and the empty-article prior gives it 32%. In both markets the articles move each outcome by only 1–4 points, and in both the disagreeing favourite is the first option listed.

We could not determine whether this is a position or label bias of the reader, or a genuine gap between a model trained on past Fed cycles and this year’s market. It is a clear direction to explore: shuffle the option order, reword the labels, compare models. Either answer is useful: a debiased reader, or a measurable “base-rate vs market” gap.

Ranking and movement are different questions. On the same Fed markets the ranking is off, but the *change* in the news read follows the market closely: weekly co-movement +0.35 to +0.44 on the September and October meetings and the number of 2026 cuts. Through August–September the news tilt toward a hike rose as the market’s P(hike) went from 0.25 to 0.99.

4.3 Does the news lead the price?

Pooled over every market, we test whether the news part of the implied odds explains the next h days’ change in the price, beyond the market’s own reversion toward its recent average. Rows are non-overlapping, outcomes priced 2–98%. The null permutes the news term across dates within each outcome (500 draws), and OLS t uses errors clustered by market-day. All three beat every permutation (7 days: $p = 0.004$). The permutation null is above zero because an outcome’s news term has a persistent level that the reversion term also sees, so

the excess over the null (+0.04 to +0.05) is the honest effect size. Pooled, it is a small edge. It is larger where the news tracks best:

- by structure: ordered markets (ranges and counts: +0.111 at 1 day, +0.184 at 3 days) and by-date markets (+0.065);
- by category: economy & central banks (+0.094) and tech, AI & companies (+0.105).

It is weakest on binary yes/no markets (+0.037). It grew through the year: no single month before July is significant, and September is (Appendix D, Table D.3). Dropping the 10 markets with the strongest own result leaves +0.071, and leaving out any one category leaves +0.057 to +0.084.

We did not test a trading strategy: no spreads, fees, sizing or execution. The question of whether this signal, or a per-market version of it, can be traded is open.

4.4 Who writes the queries, and how broadly

On the 12 markets we had built by hand, we compared three versions over the same days: ours, Claude with strict guidelines (the filter must find mostly relevant articles), and Claude with recall-first guidelines (several hundred articles a day is fine):

We bootstrapped dates, using the same resample for all three versions:

- strict against hand: +0.02 at 1 day (strict ahead in 76% of resamples) and +0.07 at 3 days (96%);
- strict against recall-first: +0.05 at both horizons (100% and 95%).

Reading more articles made the forward signal worse: off-topic articles push the read around.

setting (news smoothing, transform)	median r	null median	markets above 0.5 (null)	above 0.7	negative
EMA 14 days, level	0.165	−0.012	24.8% (1.0%)	11.9%	34.6%
average 30 days, level	0.182	0.001	23.8% (1.0%)	10.9%	35.6%
EMA 7 days, level	0.181	0.022	22.8% (1.0%)	8.9%	29.7%
EMA 7 days, z 90 days	0.215	0.012	16.8% (0%)	2.2%	21.4%
EMA 14 days, z 60 days	0.184	0.009	14.8% (0%)	3.0%	22.8%
EMA 14 days, z 30 days	0.132	−0.001	5.9% (0%)	0%	24.8%

what ranks the outcomes	same favourite as the market	pairs in the market's order
chance	41.0%	50%
the model, no article	54.1%	63.0%
the model, reading the articles	59.3%	67.1%
the articles' effect alone (no-article answer subtracted)	47.8%	55.0%

On 20 further markets with no hand version, the strict drafts scored +0.052 at 1 day (null 95% +0.028) and +0.081 at 3 days (+0.058). Their weak spot was event-risk questions (“will Putin leave office”, “will China invade Taiwan”). Claude filtered on the outcome itself (“Putin resigns”), which almost never appears until it happens: 0.6–7 articles a day.

We added a guideline: filter on what drives the risk (escalation, drills, health, succession, coalition votes), and have the relevance judge count driver articles. The redrafts read 6–280 articles a day. Their weekly co-movement went from −0.13 to +0.17 in the same week and from −0.03 to +0.30 in the next.

A second round did not help, as a procedure. On 13 markets, Claude saw its first draft’s Jan–Jun results and wrote a second version. We kept, per market, the better one on Jan–Jun and compared on Jul–Sep. Version 2 gained +0.000 / +0.005 at 1 / 3 days, and the per-market picks +0.002 / +0.008 (Appendix D, Table D.4).

Individual revisions helped a lot (crude oil all-time high: 3-day +0.20 → +0.40) or hurt a lot (a Supreme Court retirement: +0.28 → −0.23). Six months of one market could not tell them apart. Lessons pooled across many markets, fed back into the guidelines, are the more promising loop.

5. What to investigate next

- **Why some markets track and others don’t.** Coverage volume, outlet mix, how often the underlying facts change, whether the market moves on non-English or non-news information, and the share of off-topic articles each filter lets in. The per-market results are published, so anyone can try.
- **The negative tail.** About 30% of markets correlate negatively. Each is either a reader or filter failure we can fix, or a real case of a market moving against its news.
- **Prior vs market.** Where the empty-article answer disagrees with the market, is it a reader bias (option order, label wording) or a measurable base-rate gap? Shuffled options and a second model would settle much of it.
- **Per-market windows and horizons.** The best setting differs by market. A market-specific choice made out of sample, with a held-out period, is the honest way to use that.

- **Trading.** We did not try. A test that accounts for spreads, fees and capacity, on the live feed, is the obvious follow-up, and we would like to see one.
- **Live, out-of-sample results.** The feed has published every hour since 2026-10-03; we will report on it.

6. What did not work

- **Asking whether each outcome will go up.** Our first read (40 markets, 2.9M reads, all of 2026) asked per outcome “the probability that {outcome} is likely to increase / decrease”. It read tone, not content: the outcomes of one market correlated 0.87–0.90 article by article. Same-day correlation with the market was +0.003, and the next-day sign right 50.3%. Reading the outcomes against each other, relative to an empty article, is what changed the results.
- **One prompt per article.** Three short prompts per article cost little more than one, because the article is the shared prefix (43k vs 18.5k billed tokens/s). No single prompt wins everywhere: mc alone ranks best (level +0.19 against +0.17 for the average), and the average co-moves best by a hair.
- **Filtering on the full article.** The filter originally ran on the full text while the model read the first 260 words. An IPO mentioned in paragraph 8 routed the article to the IPO market, but the model never saw that paragraph. Filtering the same 260 words took the Anthropic IPO market from 25% to 45% of sampled articles judged relevant.
- **Broad filters** (4.4).
- **Faster reads.** With the articles’ 15-minute timestamps and 3-hour prices, news in the last 3–24 hours correlated at most +0.03 with the move in the same window. Daily resolution is where the relationship shows.
- **Summarising the day first.** Rendering a market-day’s reads as text and asking the model once gave the best weekly dynamics (location +0.33, hazard +0.46), but ranked outcomes no better than chance.
- **Picking the better query per market from its own history** (4.4).

7. Threats to validity

- **The backtest is not out of sample in time.** Every query, hand-built or drafted, was written on 2026-10-01 to 10-

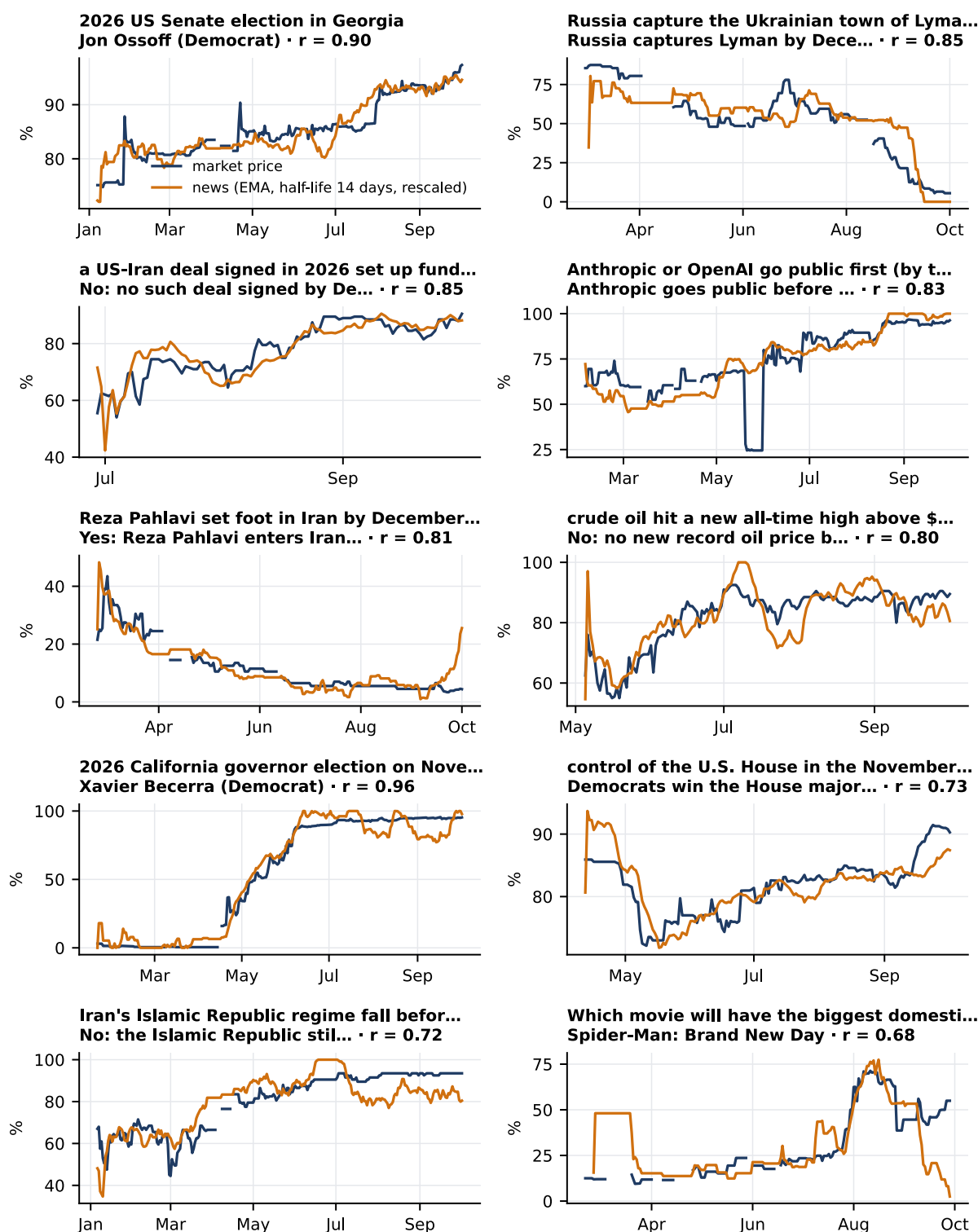


Figure 4: The ten markets (one per event, 12+ weeks) whose news read follows the price most closely. Blue: the market price of the outcome that moved most. Orange: the news read for that outcome (EMA, half-life 14 days), rescaled linearly to the price's mean and spread over the market's history; the rescaling does not change the correlation (r , computed on log-odds). The news line is the articles alone: no price information goes into it.

03, by people or a model that could see which markets exist now. In Claude's case, its training data runs to June 2026. A filter naming someone who only became relevant in August encodes hindsight. Markets are also chosen by today's volume, so markets that faded are missing. Both bias toward finding a relationship. The feed's history files carry `backfill: true` for this reason, and the live feed

is the clean test.

- **Correlations of levels need a null**, which is why Section 4.1 reports one for every setting. The cross-market null controls for "two trending lines", but not for a news series and a price that both respond to the same calendar (an approaching deadline). Within-market tests (4.3) do control

Do the articles rank the outcomes as the market does?

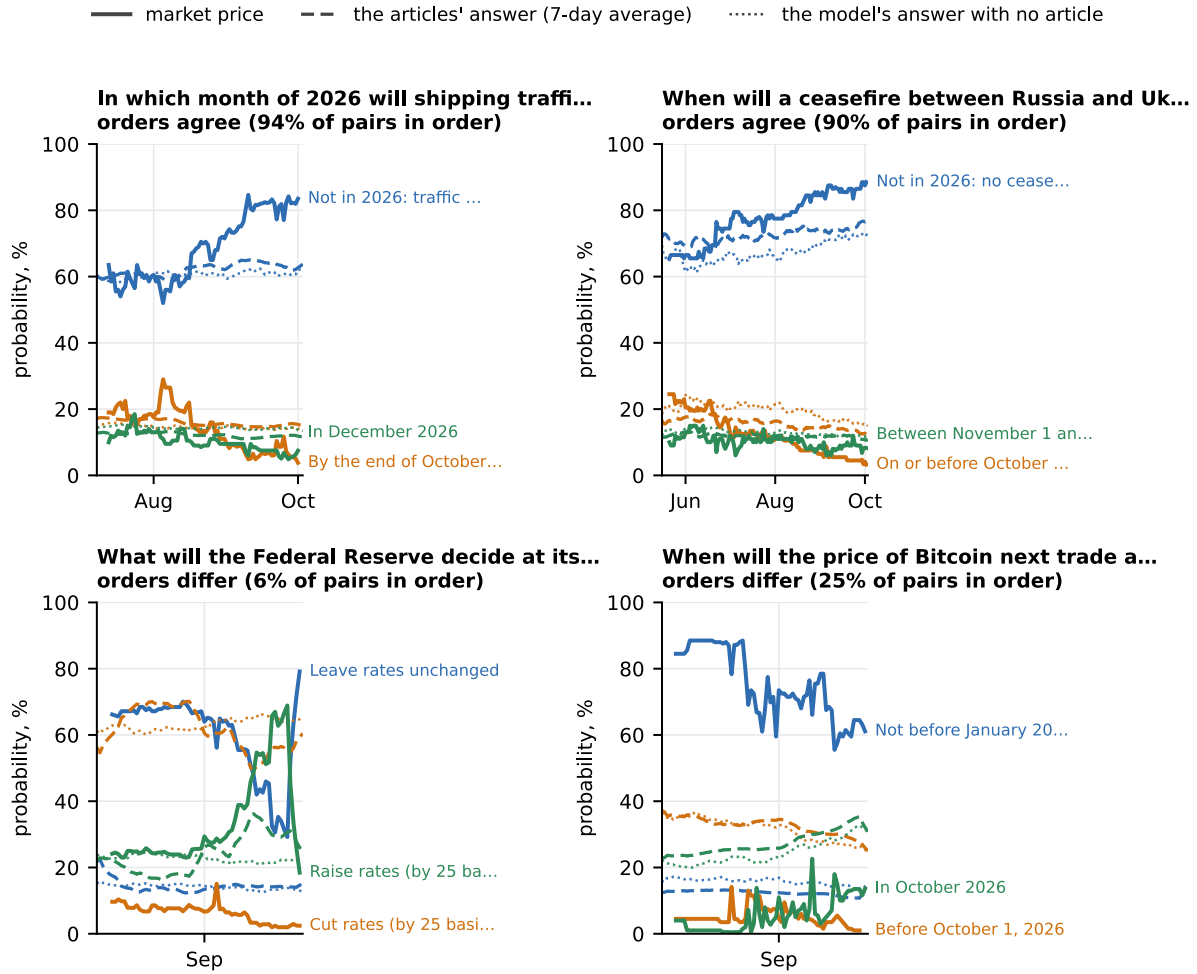


Figure 5: Do the articles rank the outcomes as the market does? One panel per market, its main outcomes in one colour each: solid, the market's probability; dashed, the articles' answer (the mean multiple-choice probability over the day's articles, 7-day average); dotted, the model's answer to the same question with no article. Top: two markets whose orders agree most (90–94% of outcome pairs in the market's order). Bottom: two whose orders differ most; there the dashed and dotted lines nearly coincide, so the disagreement is the model's prior more than the articles.

horizon	rows (markets)	partial r	null mean	null 95%	excess	t (clustered)	sign right
1 day	58,418 (195)	+0.073	+0.031	+0.040	+0.042	9.7	50.7%
3 days	19,055 (190)	+0.100	+0.047	+0.064	+0.054	6.0	51.1%
7 days	8,229 (181)	+0.084	+0.048	+0.074	+0.036	4.9	51.3%

for that.

- **Many cuts.** We report 28 smoothing settings, three horizons, seven categories, four structures and nine months. Some results below $p = 0.05$ are expected by chance among them. The pooled tests and the distribution in Figure 3 are the results we would defend.
- **Coverage is not neutral.** English only, and our outlet selection decides what “the news” is. US politics is over-represented, and so are outlets that publish a lot.
- **Thin markets bounce.** Daily changes on thin order books reverse (next-day autocorrelation -0.2 to -0.4 on Poly-market US). This is why the forward tests control for reversion and Section 4.1 requires continuous prices.
- **The reader has priors.** Section 4.2 shows how much of a

raw answer is the prior.

8. The live feed

Every hour (about :22 past the hour, UTC), for every market of the current top 200:

- **current state per outcome:** price, news-implied odds, flag, the news read and its 1-day / 7-day change in its own standard deviations, articles today against the norm, news rank vs market rank;
- **change events:** news moved; news moved while the price held; alignment changed; ranking changed; unusual attention. Each event carries a severity calibrated on the 2026 replay, and there are three subscription tiers: major

	hand-built	Claude, strict	Claude, recall-first
articles a day per market	174	14	77
forward, 1 day (null 95%)	+0.111 (0.088)	+0.130 (0.075)	+0.080 (0.062)
forward, 3 days (null 95%)	+0.110 (0.121)	+0.186 (0.106)	+0.137 (0.108)

(~ 1 a week per market), normal (~ 1.4 a day), all (~ 2.5 a day). Messages are rule-based templates with numbers (“Leave rates unchanged: news up today (4.5σ , 129 articles); market up 7 pts today ($72\% \rightarrow 79\%$)”), with links to the articles that drove the move. No generated text;

- **files and lists:** daily, weekly and monthly files and top lists, a sortable list of all markets, and per-market history.

Every file except two small pointers (`latest.json`, `manifest.json`) is immutable, and final days are never rewritten. When we revise a market’s query, it gets a new version with a new full history, and the old version stays published. The data, its documentation and a free API key are at thinkingtext.com/polymarket-news-index.

Not published: the per-market queries and filters, and the drafting guidelines in full. Appendix B shows five markets exactly as configured, so the method can be reproduced.

9. Cost

- **Reading:** the final read of 2026-10-01 took 6.7M billed tokens (3,019 articles, 181 markets, 25,917 article-market-prompt queries, ~ 10 minutes on two GPUs). At 250 markets that is ~ 8 – 10 M a day. Hourly provisional reads add only articles not yet read.
- **Drafting:** one Claude call (sometimes two) per new market plus a cheap relevance check, about a minute.
- **GDEL:** one download and one filtering pass a day for every outlet list of every project (22 s for six lists), shared with our other news indexes.

Appendix A. The prompts

The article (its first 260 words) is {X}; {?} is where the API reads the continuations. The null read replaces the article with N/A.

mc

```
{X}

Today is {DATE}.
Question: {Q}
{OPTIONS}
Based on the article above, the most likely answer is option{?}
```

continuations: A, B, ... one per open outcome.

yn (one prompt per outcome)

```
{X}

Today is {DATE}.
Question: {Q}
Based on the article above, is the answer "{0}"? Answer:{?}
```

continuations: Yes, No.

dir (one prompt per outcome)

```
{X}

Today is {DATE}.
Question: {Q}
Summary of the key facts in the article above, from beginning to end, and their consequence: the chance that t
he answer is "{0}" is likely to{?}
```

continuations: increase, decrease.

As the model reads it (mc, the September Fed meeting, an article of 2026-08-20):

```
<the article's first 260 words>

Today is August 20, 2026.
Question: What will the Federal Reserve decide at its September 15-16, 2026 meeting?
A) Cut rates by 50 basis points or more
B) Cut rates by 25 basis points
C) Leave rates unchanged
D) Raise rates by 25 basis points
E) Raise rates by 50 basis points or more
Based on the article above, the most likely answer is option
```

Appendix B. Five markets, exactly as configured

Filters: an OR of rules; a rule is an AND of groups; a group an OR of terms; = = case-sensitive; *sentence* = all groups in one sentence. r is the correlation of Section 4.1 (EMA 14 days, level).

B.1 How many Fed rate cuts in 2026? (polymarket.com, Claude-drafted; 38 weeks, 136 articles/day; weekly co-movement +0.44, next week +0.28) - Outcomes: No cuts = $P(\text{no cuts})$ · One 25 bp cut = $P(1 \text{ cut})$ · Two or more = $\sum P(2 \dots 12+ \text{ cuts})$ - Filter (*sentence*): [Federal Reserve, =FOMC, =Fed, Jerome Powell, Powell, Kevin Warsh, Warsh] AND [rate, rates, rate cut, rate cuts, rate hike, interest rates, basis points, monetary policy, federal funds, easing, rate decision]

B.2 Which party wins the US House on November 3, 2026? (Polymarket US, Claude-drafted; 26 weeks, 30 articles/day; $r = 0.73$; same favourite 80% of days) - Outcomes: Democrats win the House majority · Republicans keep the House majority (one contract each) - Filter: [midterm, midterms, midterm election(s), November election, Election Day] AND [House majority, control of the House, flip the House, keep the House, House seats, control of Congress,

House Democrats, House Republicans, ...]; or [House majority, control of the House, generic ballot, competitive districts, ...] AND [poll(s), forecast, Cook Political Report, ...]

B.3 When will Anthropic confirm an IPO? (Polymarket US, hand-built; 17 weeks, 107 articles/day; forward 1 day +0.17, 3 days +0.16) - Outcomes: by Oct 31 = $P(\text{by Oct 31}) \cdot \text{in November} = P(\text{by Nov 30}) - P(\text{by Oct 31}) \cdot \text{in December} \cdot \text{Jan-Jun 2027} \cdot \text{after June 2027 or never} = 1 - P(\text{by Jun 30 2027})$ - Filter (sentence): [=Anthropic] AND [IPO, initial public offering, go public, going public, public offering, S-1, stock market debut, IPO filing, ...]

B.4 Will Iran's Islamic Republic fall before the end of 2026? (polymarket.com, Claude-drafted; 39 weeks, 155 articles/day; $r = 0.72$; same favourite 98% of days) - Outcomes: Yes = $P(\text{yes}) \cdot \text{No} = 1 - P(\text{yes})$ - Filter: [Iran, Iranian, Tehran, Islamic Republic, IRGC, Revolutionary Guards] AND [regime change, regime collapse, overthrow, uprising, coup, civil war, defections, transitional government, anti-government protests, ...]; or [Khamenei, Mojtaba Khamenei, Ali Khamenei] AND [succession, successor, health, death, assassination, Assembly of Experts, ouster, ...]; or [Reza Pahlavi] AND [Iran, regime, opposition, ...]

B.5 Will China launch a military offensive on Taiwan by December 31, 2026? (polymarket.com, Claude-drafted after the event-risk guideline; 39 weeks, 12 articles/day; weekly co-movement -0.13 , next week $+0.26$) - Outcomes: Yes = $P(\text{yes}) \cdot \text{No} = 1 - P(\text{yes})$ - Filter (sentence): [China, Chinese, Beijing, =PLA, People's Liberation Army, Eastern Theater Command, China Coast Guard] AND [Taiwan, Taiwanese, Taipei, Kinmen, Matsu, Penghu, Taiwan Strait] AND [drills, military exercises, war games, live-fire, warships, blockade, incursion(s), amphibious, encirclement, invasion, combat readiness, ...]; or [Taiwan Strait] AND [drills, warships, blockade, ...]

Appendix C. Reproducing the numbers

The published data (thinkingtext.com/polymarket-news-index) carries, per market, outcome and day, everything this paper uses: - the price; - the raw news read and its smoothings (EMA 7, 14, 30 days), plus the chart line of Figures 1 and 4 (the 14-day EMA rescaled onto the price's range with past data only); - the news-implied odds and flag; - the article counts; - the articles' answer share, the model's no-article answer share, and the ranks of all three; - per market: the correlations of Section 4.1 (level and 90-day z-score), weekly co-movement, the 3-day forward partial and the ordering statistics of Section 4.2.

Sections 4.1 to 4.4 can be recomputed from it.

Appendix D. More results

Table D.1 Description by category (ordering on market-days with 3+ articles; co-movement is the per-market mean).

category	markets	articles/day (median)	same favourite (chance)	pairs in order	weekly co-movement
Geopolitics & conflict	87	20	77.1% (44.5%)	80.2%	+0.065
Elections	65	8	47.6% (40.0%)	56.9%	+0.114
Tech, AI & companies	25	8	51.6% (34.3%)	65.8%	+0.103
Economy & central banks	17	243	34.1% (35.2%)	49.6%	+0.294
Crypto	9	12	60.6% (43.6%)	65.0%	+0.169
Culture & media	5	13	31.9% (32.7%)	55.8%	+0.107
Other	13	14	65.6% (37.3%)	78.7%	+0.012

Table D.2 Forward test by market structure, category and query source (partial r , permutation null mean in brackets).

	1 day partial (null mean)	p	3 days partial (null mean)	p
ordered (ranges, counts)	+0.111 (+0.060)	<0.01	+0.184 (+0.091)	0.01
by-date	+0.065 (+0.015)	<0.01	+0.067 (+0.023)	<0.01
contenders	+0.049 (+0.016)	<0.01	+0.066 (+0.030)	0.02
binary	+0.037 (+0.013)	0.03	+0.032 (+0.024)	0.38
Economy & central banks	+0.094 (+0.028)	<0.01	+0.107 (+0.031)	0.02
Tech, AI & companies	+0.105 (+0.059)	<0.01	+0.185 (+0.101)	0.01

	1 day partial (null mean)	p	3 days partial (null mean)	p
Elections	+0.051 (+0.014)	<0.01	+0.062 (+0.029)	0.03
Geopolitics & conflict	+0.038 (+0.009)	<0.01	+0.043 (+0.016)	0.05
hand-built queries (19 markets)	+0.121 (+0.062)	<0.01	+0.188 (+0.090)	<0.01
Claude-drafted (176 markets)	+0.044 (+0.016)	<0.01	+0.057 (+0.026)	0.01

The hand-built markets are the biggest markets with the most coverage (median 156 articles a day against 11), so the last two rows compare markets as much as authors. Section 4.4 compares authors on the same markets.

Table D.3 Forward test at 1 day, by month.

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
partial r	0.021	0.012	0.039	0.026	0.049	0.099	0.061	0.074	0.086
null mean	0.006	0.023	0.030	0.042	0.034	0.070	0.044	0.050	0.054
p	0.15	0.78	0.25	0.78	0.17	0.06	0.03	0.05	<0.01

Table D.4 Feedback loop: second versions written from Jan–Jun results, compared on Jul–Sep (13 markets).

held out, Jul–Sep	version 1	version 2	per-market pick
forward 1 day	0.041	0.042	0.043
forward 3 days	0.059	0.063	0.067
gain over v1, $P(> 0)$		0.50 / 0.55	0.68 / 0.84