

Quarterly Consensus Watch #1

October 8, 2026 · MarketMania Research · Quarterly series · window Jul 15-Sep 30, 2026 UTC

RESEARCH SNAPSHOT			
50,707 LOO-scoreable calls	7 models (lineage spliced: opus, grok, qwen)	8,899 thin/tied excluded	window Q3 2026 · Jul 15-Sep 30 · partial
method = leave-one-out majority, 3+ directional peers	FH gate 1h / 4h / 1d	base field hit 45.6%	cutoff Thu Oct 1, 2026, 16:00 UTC
market: BTC net +28.56% (prior n/a)	ann vol 38.9% (was n/a)	TOP5 volume \$173.94B, n/a q/q	pairwise corr 0.77 (was n/a)

KEY FINDING · OBSERVATION (one quarterly window)

Herding ran at 98.5% of leave-one-out scoreable calls across the quarter and the pooled lift came in at -7.0pp: agree-hit 45.1% against 52.0% on disagreement, on a disagree bucket of n=784.

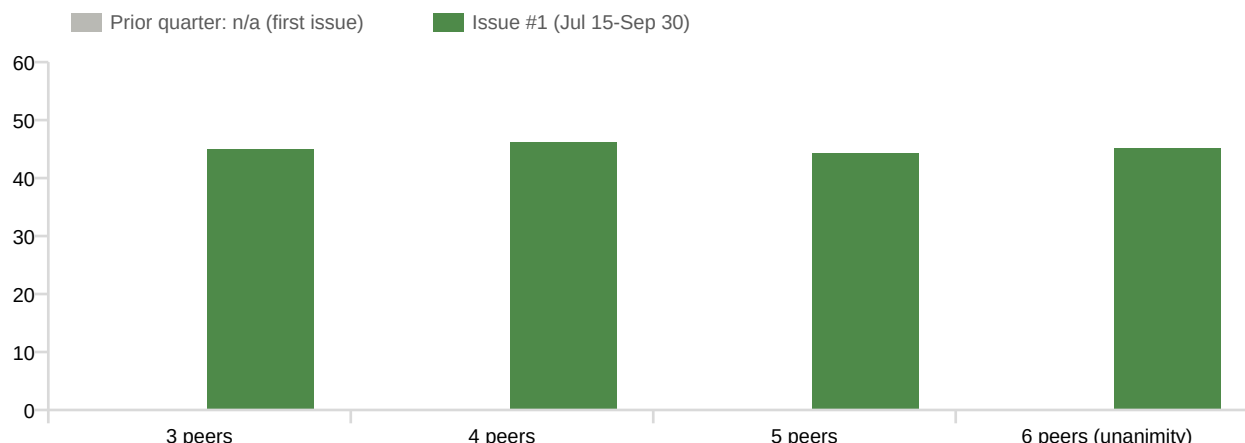
MONTHLY EVIDENCE -> QUARTERLY VERDICT (rule below)

Metric	M1 Jul 15-31	M2 Aug 1-31	M3 Sep 1-30	Quarter	Verdict
LOO lift, pp	-10.6	+1.9	-9.7	-7.0	Confirmed (2/3)
Herd share >= 95%	97.4%	98.7%	99.2%	98.5%	Confirmed (3/3)
Unanimity hit vs field hit	+0.9	-1.2	-1.2	-0.5	Confirmed (2/3)

Verdict rule: each row states one reading of the quarter; a month agrees when it shows the same reading (same sign, or the same name). Confirmed = 2 or 3 of the 3 months agree with the quarter; Rejected = 0 or 1. M1 holds Jul 15-31 only: the platform's first forecast slot is Jul 15 18:01 UTC. Monthly cells come from the frozen monthly packs (7-line roster, lineage spliced); the quarter is the pooled Jul 15-Sep 30 pull. The M1 unanimity cell uses the same 6-peer tier as the other cells (45.9% vs the 45.0% field, +0.9); Monthly #1 printed July's unanimity by the wider herd rule (45.6%, +0.6), the same sign.

TL;DR

- **OBSERVATION -- 98.5% of the field's leave-one-out calls ran with the herd over the quarter.** Of 50,707 LOO-scoreable directional calls, 49,923 (98.5%) sided with peers' majority; agree-hit was 45.1% [44.6%, 45.5%] vs. 52.0% [48.5%, 55.5%] for the 784 disagreeers, a pooled lift of -7.0pp. Descriptive, one quarter, not a claim of a durable edge.
- **The margin curve read 44.9% / 46.1% / 44.3% / 45.1% from 3 peers to unanimity.** 1 of the 4 tiers sit above the 45.6% field base. By month the same four tiers read Jul 15-31 42.7% / 43.5% / 43.1% / 45.9%, Aug 1-31 45.9% / 45.1% / 45.8% / 45.1%, Sep 1-30 45.9% / 49.7% / 43.8% / 44.3%; there is no prior quarter.
- **Month by month the lift changed sign 2 times.** The three published monthly windows ran Jul 15-31 -10.6pp, Aug 1-31 +1.9pp, Sep 1-30 -9.7pp; the quarter pools at -7.0pp, and there is no prior quarter. The quarter is one pooled pull, not the mean of the months: the monthly range is -10.6pp to +1.9pp.
- **5,083 cells had 6+ directional models on one side across the quarter.** The 5 worst sat at 71.5-73.0 mean confidence and resolved 0.0%-0.0%; by month 1,446 / 1,672 / 1,965 such cells (Jul 15-31 / Aug 1-31 / Sep 1-30).



Hit rate (%) by number of agreeing LOO peers over the whole quarter. Quarterly n: 3 peers 5,285 / 4 peers 7,424 / 5 peers 12,082 / 6 peers 24,770. Field base this quarter 45.6% (Jul 15-31 45.0%, Aug 1-31 46.3%, Sep 1-30 45.5%). First quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026), so the chart carries no gray series and every prior cell in this report reads n/a; a month inside the quarter is not a prior window.

Why it matters

MarketMania publishes forecasts from 7 frontier model lines into the same slots every hour; a natural question for anyone reading the feed is whether a model's call is worth more when it agrees with what everyone else is saying, or whether that agreement is just everyone making the same mistake together. The three monthly windows inside this quarter answered it with lifts of Jul 15-31 -10.6pp, Aug 1-31 +1.9pp, Sep 1-30 -9.7pp, and the pooled quarter reads -7.0pp. This quarterly issue asks the same question at quarterly n, across three months the pack labels as 1 calm and 2 volatile, using the same leave-one-out design built to avoid comparing a model to an index that includes its own vote.

How to read this

For every model M and every forecast it made, we rebuild that slot's consensus using only the OTHER active models in the same symbol / forecast-horizon (FH) / timeframe (TF) cell -- M's own call never counts toward its own consensus. M is scored agree if its side (long or short) matches the strict majority of those peers, and disagree if it sits alone against that majority; a cell needs at least 3 directional peers to count at all, and tied peer splits are excluded rather than forced either way. LOO peers are genuinely external to the model being scored, so the comparison below is not comparing a model to itself. 3 model lines are lineage splices this quarter (claude-opus-5 (incl. 4.8-era, Jul 15-30); grok-4.6 (incl. 4.5-era, Jul 15-Aug 24); qwen-3.8-max (incl. 3.7-era, Jul 15-Aug 5)), and a spliced line never counts as two peers in a cell. The quarter is one pull at one cutoff (Thu Oct 1, 2026, 16:00 UTC); the three monthly blocks quoted alongside it are the published monthly windows at their own frozen cutoffs.

Pooled result: agree vs. disagree with the LOO consensus

Group	n	Hit rate	Wilson 95% CI	Lift, pp
Agree with LOO peers	49,923	45.1%	[44.6%, 45.5%]	-7.0 (vs. disagree)
Disagree with LOO peers	784	52.0%	[48.5%, 55.5%]	ref.

The disagree bucket is n=784 over the whole quarter; the two Wilson intervals do not overlap. The prior window reads n/a because this is the first quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026). By month (Jul 15-31 / Aug 1-31 / Sep 1-30): agree 44.6% / 45.5% / 45.1%, disagree 55.3% / 43.6% / 54.7%, lift -10.6pp / +1.9pp / -9.7pp, on n=436 / 211 / 137 disagreeing calls. These are dependent observations pooled across the whole quarter: one calm month and two volatile months.

By forecast horizon / timeframe

The pooled result above mixes 6 different FH/TF cells. Broken out:

FH / TF	Agree n	Agree hit	Dis. n	Dis. hit	Lift, pp	Excluded
1h / 1h	28,469	46.2%	510	54.1%	-7.9	5,163
1h / 15m	2,706	45.8%	40	42.5%	+3.3	592
4h / 4h	7,304	44.9%	97	47.4%	-2.5	1,397
4h / 1h	8,503	43.7%	109	51.4%	-7.7	1,316
1d / 1d	1,411	38.7%	19	47.4%	-8.7	247
1d / 4h	1,530	37.5%	9	44.4%	-7.0	184

"Excluded" = thin (fewer than 3 directional peers) or tied peer splits in that cell; not folded into agree or disagree. "Dis." = disagree. Cells whose disagree bucket is under N=10 are insufficient and are never used to rank anything. 1h / 15m holds rows of Jul 15-31 only.

Margin: does more agreement mean a safer call?

For every LOO-scored forecast that sided with its peers' majority (the agree bucket, n=49,923) we also count how many peers were on that majority side -- from the bare minimum of 3 up to the full peer set agreeing (6 of 6, since 7 model lines were active and LOO always drops one); the 784 disagreeing calls sit in no tier, so the tiers sum to the agree bucket, not to the 50,707 scoreable calls. The quarter-over-quarter chart is on page 2; the full per-tier numbers: 3 peers 44.9% (n=5,285), 4 peers 46.1% (n=7,424), 5 peers 44.3% (n=12,082), 6 peers 45.1% (n=24,770). Against the 45.6% field base: 1 tier above, 3 tiers below; there is no prior quarter. The month-by-month table further down prints each tier for each of the three months.

Daily-horizon herds: the thinnest cells in the report

The two daily cells stay the report's thinnest even at quarterly n. 1d/1d unanimity hit 35.8% (n=758) and 1d/4h unanimity 37.3% (n=879) against the 45.6% field base; the whole 1d/1d agree bucket ran 38.7% (n=1,411) -- not the highest of the six FH/TF cells -- while 1d/4h ran 37.5% (n=1,530). Both daily buckets are an order of magnitude thinner than the hourly one (n=28,469), so they are reported, not ranked.

Herd failures: the 5 worst unanimous misses

When the whole field agreed, and was wrong. This quarter 5,083 slot / symbol / FH / TF cells had 6 or more directional models sitting on the same side.

Slot (UTC)	Symbol	FH / TF	Side	Models	Mean conf	Hit rate
Jul 21, 14:01	BTC	1h / 15m	long	6	73.0	0.0%
Jul 20, 00:01	SOL	1h / 1h	long	6	72.5	0.0%
Jul 21, 14:01	BTC	1h / 1h	long	11	72.5	0.0%
Aug 4, 19:01	BTC	1h / 1h	long	7	71.7	0.0%
Jul 27, 08:01	ETH	4h / 4h	long	6	71.5	0.0%

The 5 worst cells were long swarms at 71.5-73.0 mean confidence and resolved 0.0%-0.0%. 'Sideways' was 51.4% of mature forecasts (63,164 of 122,770) this quarter. 'Models' counts the directional forecast rows in the cell, as the pack does; the Jul 21, 14:01 BTC 1h/1h cell carries 11 rows against a 7-line roster, so at least one line sits in it more than once.

Per-model: agree vs. disagree

Per-model disagree counts run 29 to 231 over the quarter. 7 of the 7 lines clear the N=10 floor (claude-fable-5 29, claude-opus-5 (spliced) 72, deepseek-v4-pro 119, gemini-3.1-pro 231, gpt-5.6-sol 31, grok-4.6 (spliced) 127, qwen-3.8-max (spliced) 175); every other per-model contrarian cell below is insufficient.

Model	Agree n	Agree hit	Disagree n	Disagree hit
claude-fable-5	7,444	45.2%	29	48.3%
claude-opus-5 (spliced)	7,304	45.5%	72	55.6%
deepseek-v4-pro	6,339	44.9%	119	42.9%
gemini-3.1-pro	7,208	45.9%	231	54.5%
gpt-5.6-sol	7,815	44.6%	31	51.6%
grok-4.6 (spliced)	6,222	44.8%	127	50.4%
qwen-3.8-max (spliced)	7,591	44.6%	175	55.4%

The prior window reads n/a for every line because this is the first quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026). By month the per-model agree-side hit ran 44.1% to 45.2% (Jul 15-31), 44.5% to 47.3% (Aug 1-31), 43.5% to 47.2% (Sep 1-30). At these counts a handful of flipped calls moves a per-model lift by tens of points, so models are not ranked by contrarian lift here. claude-opus-5 (spliced) = claude-opus-5 (incl. 4.8-era, Jul 15-30). grok-4.6 (spliced) = grok-4.6 (incl. 4.5-era, Jul 15-Aug 24). qwen-3.8-max (spliced) = qwen-3.8-max (incl. 3.7-era, Jul 15-Aug 5).

Month by month: how often the field agreed, and what it paid

The same leave-one-out scoring run on each of the three months of the quarter gives three monthly lifts; the pooled quarterly number covers the same days, as no day of the quarter falls outside the three months. Herd share ran 97.4% to 99.2% and the lift ran -10.6pp to +1.9pp across the three months against -7.0pp for the quarter as a whole.

Month	Scoreable	Herd share	Agree hit	Disagree hit	Lift pp	Unanimity n	Unanimity hit
M1 Jul 15-31	16,993	97.4%	44.6%	55.3%	-10.6	11,531	45.6%
M2 Aug 1-31	16,089	98.7%	45.5%	43.6%	+1.9	10,984	45.4%
M3 Sep 1-30	17,625	99.2%	45.1%	54.7%	-9.7	13,121	44.2%
Quarter	50,707	98.5%	45.1%	52.0%	-7.0	35,636	45.0%

Source: the quarter pull split by slot month at the single quarterly cutoff (Thu Oct 1, 2026, 16:00 UTC), so the three month rows and the Quarter row are one consistent pool. The frozen-cutoff cross-check is the published monthly windows at their own cutoffs: the recomputed lift reproduces the published lift for 3 of 3 months (M1 -10.6pp vs -10.6pp, M2 +1.9pp vs +1.9pp, M3 -9.7pp vs -9.7pp).

* = N<10 (insufficient) -- such cells are never ranked.

Margin	M1	M2	M3	Quarter
3 peers	42.7% (1,609)	45.9% (1,964)	45.9% (1,712)	44.9% (5,285)
4 peers	43.5% (2,394)	45.1% (2,630)	49.7% (2,400)	46.1% (7,424)
5 peers	43.1% (3,628)	45.8% (4,482)	43.8% (3,972)	44.3% (12,082)
6 peers (unanimity)	45.9% (8,789)	45.1% (6,664)	44.3% (9,317)	45.1% (24,770)

Cells are hit rate with n in parentheses; * = N<10 (insufficient), "n/a (0)" = no call landed in that tier that month. Months are M1 Jul 15-31, M2 Aug 1-31, M3 Sep 1-30. Same source and cutoff as the table above; the Quarter column is the pooled quarter, not the mean of the three months. Quarter-over-quarter columns would compare the prior quarter against Jul 15-Sep 30 (this issue); every prior cell reads n/a because this is the first quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026).

Market check: a calm July, then two volatile months

OBSERVATION -- mean |1d move| 1.54% / 1.98% / 1.91% by month (1.81% over the quarter), BTC realized vol 33.5% / 36.9% / 34.9% (35.1%; ann., hourly); field directional accuracy 45.0% / 46.3% / 45.5% (45.6%), sim win-rate 31.9% / 35.8% / 38.4% (35.4%), field sim PnL -\$3095.02 / -\$1054.79 / -\$2554.15 (-\$6703.95) (Jul 15-31 / Aug 1-31 / Sep 1-30, quarter in parentheses; descriptive, three windows, not a claim).

Robustness: raw price-sign accuracy 46.8% / 46.0% / 48.4% by month (47.1% over the quarter) against 45.0% / 46.3% / 45.5% (45.6%) under the trade-based hit rule; month to month the two moved in the same direction in 0 of 2 steps.

Measure	Jul 15-31	Aug 1-31	Sep 1-30	Quarter
Mean 1d move (5 assets)	1.54%	1.98%	1.91%	1.81%
BTC realized vol (ann., hourly)	33.5%	36.9%	34.9%	35.1%
Field directional accuracy	45.0%	46.3%	45.5%	45.6%
Models improving hit-rate	--	5 of 7	1 of 7	--
Raw price-sign accuracy	46.8%	46.0%	48.4%	47.1%
Field sim win-rate	31.9%	35.8%	38.4%	35.4%
Field sim net PnL	-\$3095.02	-\$1054.79	-\$2554.15	-\$6703.95

There is no prior quarter, so this slice prints the three months and the quarter side by side instead of a prior-against-current pair; each month is its own pack's window at its own frozen cutoff and the Quarter column is the quarter pull. Descriptive, three windows; the hit rule is methodology v1.1 (trade-based) and the raw price-sign row is the robustness check (hourly closes at :00 against slots at :01). "Models improving" compares a month with the month before it (Aug 1-31 against Jul 15-31; Sep 1-30 against Aug 1-31). BTC realized vol here is the hourly-return series from the alive slice; the snapshot row's figure is the audited daily-candle estimate -- different estimators, both reported as measured. Lineage splices in this window: claude-opus-5 (incl. 4.8-era, Jul 15-30); grok-4.6 (incl. 4.5-era, Jul 15-Aug 24); qwen-3.8-max (incl. 3.7-era, Jul 15-Aug 5). 24,922 rows were remapped inside the quarter (19,532 / 5,390 / 0 in Jul 15-31 / Aug 1-31 / Sep 1-30).

Practical implications

- 1 of the 4 margin tiers sit above the 45.6% field base this quarter; the four tier rates span 44.3% to 46.1%, a spread of +1.7pp, and the three months inside the quarter ran lifts of -10.6pp to +1.9pp against a pooled quarterly lift of -7.0pp.
- The pooled quarterly lift is -7.0pp while the three months ran -10.6pp to +1.9pp; there is no prior quarter to compare against.
- The field base read 45.0% / 46.3% / 45.5% by month (45.6% over the quarter), and every consensus tier is measured against that base.

Limitations

- No prior quarter: this is the first quarterly issue (platform data begin Jul 15, 2026), so every "prior" cell reads n/a and the page-2 chart carries no gray series; a month inside the quarter is not a prior window. Partial quarter: the window is the calendar quarter Jul 1-Sep 30; the platform's first forecast slot is Jul 15 18:01 UTC, so 78 of 92 days carry data. Asset universe: the quarter's rows also carry 1,592 directional calls on AVAX (414), DOGE (347), LINK (436), PEPE (394) and TON (1), all in Jul 15-31; they are counted in the field totals but not ranked in the asset tables.
- Two cutoffs live in this issue. The quarter is frozen at Thu Oct 1, 2026, 16:00 UTC; the three monthly blocks are frozen at their own published cutoffs (M1 Jul 15-31, M2 Aug 1-31, M3 Sep 1-30). A forecast still pending at its month's cutoff is immature in the monthly pack and mature in the quarter pull, so the quarter need not equal the sum of the months; here the two agree (59,606 vs 59,606 directional; no day of the quarter falls outside the three months): the quarter pull split by slot month gives 19,541 / 19,383 / 20,682, the same as the frozen monthly packs.
- Observations are not independent: the same models watch overlapping symbol / FH / TF cells hour after hour for 78 days. Wilson intervals here are descriptive, not inferential, and pooling a quarter does not make them inferential.
- The pooled lift rests on 784 disagreeing calls; per-model counts run 29-231. 8,899 calls (14.9% of the 59,606-call mature directional universe) were thin or tied and are excluded from every consensus table.

Correlational, not causal: LOO removes self-match bias, not the confound that an easy slot lifts agreement and hit together.

- This window ran 24 trend days of 77 with a BTC daily move (Jul 1-15 carry none), and the pack's own volatility rule splits the three months of the quarter as M1 calm, M2 volatile, M3 volatile; every month-to-month comparison inside the quarter is a comparison across regimes as well as across windows. Series density and lineage notes (quarterly #1): (1) rolling-1w forecasts emit daily since Aug 22, 2026 and rolling-1M daily since Aug 23, 2026, so across this 92-day window daily coverage is PARTIAL as measured -- the 1w series has slots on 55 of 92 days and the 1M series on 41 of 92 days; both sit outside this report's FH gate (1h/4h/1d) and appear only in the exclusions counter (1w 4,001, 1M 2,787). (2) 3 model lines are lineage splices inside this window: claude-opus-5 (incl. 4.8-era, Jul 15-30); grok-4.6 (incl. 4.5-era, Jul 15-Aug 24); qwen-3.8-max (incl. 3.7-era, Jul 15-Aug 5). (3) Every per-model aggregate in this issue sees the merged head id; the raw id is kept in the pack, and 24,922 rows were remapped inside the quarter (19,532 / 5,390 / 0 in Jul 15-31 / Aug 1-31 / Sep 1-30).
- Market-state row is computed on a single-exchange (binance) daily candle series, as in the weekly series after the audit that found the raw table mixes two exchanges. The quarter block covers 78 days (Jul 15-Sep 30) of the 92-day calendar quarter, and there is no prior-quarter block, so the TOP5 volume figure is a 78-day sum with no quarter-over-quarter change.
- Research-to-date counter: directional forecasts resolved inside the FH gate (1h/4h/1d) since Jul 11, counted at the issue cutoff (Thu Oct 1, 2026, 16:00 UTC) -- the D2 definition pinned in the weekly series; this issue prints 59,755. The pack carries no prior-cutoff pin for a quarterly window (not applicable: there is no prior quarter, and the prior-cutoff counts are 0 by construction), so the control this issue is the monthly reproduction: the M3 block of this pack (Sep 1-30) reproduces the published Monthly #2 on 7 of 7 checks, the M2 block (Aug 1-31, cutoff Tue Sep 1, 2026, 16:00 UTC) reproduces the published Monthly #1 on 7 of 7 checks, and the M1 block (Jul 15-31, cutoff Sat Aug 1, 2026, 16:00 UTC) reproduces the prior window of Monthly #1 on 7 of 7 checks. Counter deltas against pre-wave-3 issues remain definitional.

Counters & lineage

Counter	Value	Counter	Value
Forecasts total	133,039	Mature (scored pool)	122,770
OK in gate	122,919	-- of them directional	59,606
Out of gate (1w / 1M)	4,001 / 2,787	-- of them sideways	63,164
Invalid	3,332	Pending (next issue)	149
Late closes	0	Uptime, all grids	2,929 / 3,496 slots
Source file	quarterly_metrics_2026-q3.json	Report cutoff	Thu Oct 1, 2026, 16:00 UTC
Generated at (pipeline)	2026-10-07 11:01 UTC	Methodology	v1.1 (2026-08-10) · hash e66c7e8c864a2233

THIS REPORT: 50,707 LOO-scored observations. MARKETMANIA RESEARCH TO DATE (as of cutoff Oct 1, 16:00 UTC): 59,755 resolved since Jul 11 · 7 models tracked (current line-up; earlier versions folded into their successors' lineage) · 5 assets · 5 horizons · hourly · 47 published reports.

Issue #1. Quarterly Consensus Watch is a living quarterly comparison; each issue appends one more calendar quarter of leave-one-out agreement-vs-hit data. This first issue sets the baseline on a partial quarter (Jul 15-Sep 30): herd share 98.5%, pooled lift -7.0pp (agree 45.1% vs disagree 52.0% on n=784), three monthly lifts inside the quarter running -10.6pp to +1.9pp; there is no prior quarter. 3 model lines are lineage splices this quarter (claude-opus-5 (incl. 4.8-era, Jul 15-30); grok-4.6 (incl. 4.5-era, Jul 15-Aug 24); qwen-3.8-max (incl. 3.7-era, Jul 15-Aug 5)). Engine 1.1 has powered the sandbox since Aug 18, i.e. inside this window; no cross-engine PnL comparisons are claimed.

What we're testing next

- Quarterly #2 (Q4 2026): the same three series over Oct 1-Dec 31, cut on Jan 1 plus the maturity lag, so its quarter-over-quarter comparison runs against this partial Q3.
- Quarterly Config Watch #1: the same Q3 window through the config search engine (w309), published as the fourth quarterly report of this issue.

- Quarterly Stability Index: Run 3, the next repeat-prompt stability run, is the first cross-quarter point on model stability.

Related research

Report	Direct PDF link
Quarterly Calibration #1	https://marketmania.ai/research/reports/calibration-quarterly-2026-q3.pdf
Quarterly Model Watch #1	https://marketmania.ai/research/reports/model-watch-quarterly-2026-q3.pdf
Monthly Consensus Watch #2	https://marketmania.ai/research/reports/consensus-watch-monthly-2026-09.pdf

The three quarterly reports publish together as one issue each quarter; each links straight to the others' PDF and to the last monthly issue of the same series. Direct links are the posting rule from wave 2 on.

CITE THIS REPORT

```
@misc{mm_consensus_quarterly_2026q3,
  title = {Quarterly Consensus Watch #1: does agreeing with the crowd make an LLM's market call safer?},
  author = {{MarketMania Research}},
  year = {2026}, month = {October}, day = {8},
  url = {https://marketmania.ai/research/reports/consensus-watch-quarterly-2026-q3.pdf},
  note = {Methodology v1.1; window Jul 15-Sep 30, 2026 UTC; source quarterly_metrics_2026-q3.json}
}
```