

Quarterly Calibration #1

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

RESEARCH SNAPSHOT			
59,606 directional calls scored	7 models (lineage spliced: opus, grok, qwen)	of 122,770 mature forecasts	window Q3 2026 · Jul 15-Sep 30 · partial
hit = direction rule: tp1/tp2->hit, sl->miss, expiry->sign of gross pnl	FH gate 1h / 4h / 1d	base field hit 45.6% vs 62.6 stated	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)

Over the quarter the field's overconfidence gap read +17.0pp against 62.6 stated mean confidence, the monthly gaps ran +16.3pp to +17.8pp inside it, and the field slope across the three months is -0.40 pp/month (flat).

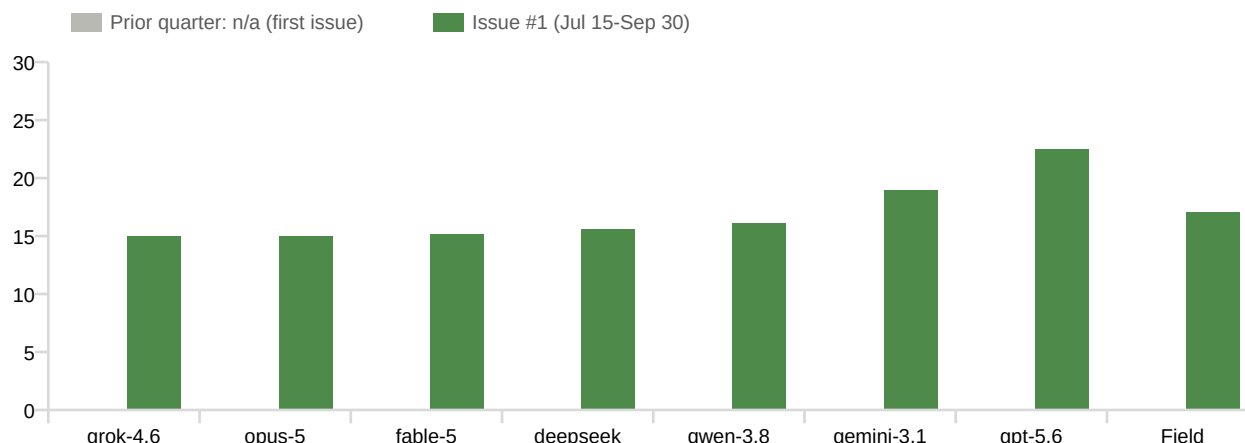
MONTHLY EVIDENCE -> QUARTERLY VERDICT (rule below)

Metric	M1 Jul 15-31	M2 Aug 1-31	M3 Sep 1-30	Quarter	Verdict
Field gap (conf - hit), pp	+17.8	+16.3	+17.0	+17.0	Confirmed (3/3)
Lowest-Brier model	grok-4.6	opus-5	qwen-3.8	grok-4.6	Rejected (1/3)
Widest-gap model	gpt-5.6	gpt-5.6	gpt-5.6	gpt-5.6	Confirmed (3/3)

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.

TL;DR

- **OBSERVATION -- The quarterly gap read +17.0pp.** Directional calls hit 45.6% against 62.6 stated mean confidence over 59,606 calls. Field Brier is 0.2803 and 0 of 7 model lines scored below 0.25, the score of an uninformative always-50% predictor. One quarter, descriptive.
- **The field gap slope across the three months is -0.40 pp/month (flat).** Across those months the field gap ran +17.8pp -> +16.3pp -> +17.0pp and the per-model slopes split 3 narrowing / 2 widening / 2 flat under the pack's own rule ($|\text{slope}| < 0.5 \text{ pp/month} = \text{flat}$). grok-4.6 (spliced) is the lowest-Brier line (Brier 0.2716); gemini-3.1-pro leads on hit-rate (47.2%).
- **The widest monthly swing belongs to qwen-3.8-max (spliced).** Its monthly gap moved over a 4.0 pp range inside one quarter, against a field range of 1.5 pp. There is no prior quarter; by month the per-model gaps ran +14.2pp to +22.1pp (Jul 15-31), +13.2pp to +21.8pp (Aug 1-31), +14.7pp to +23.5pp (Sep 1-30).
- **Trading, kept separate from every calibration table.** 0 of 7 lines finished net-positive over the quarter; net PnL ran -\$835.51 (gemini-3.1-pro, best) to -\$1141.58 (gpt-5.6-sol, worst), field -\$6703.95.



Overconfidence gap in percentage points (mean stated confidence minus hit-rate) over the whole quarter, sorted by the quarter's gap. Short names: opus-5 = claude-opus-5, fable-5 = claude-fable-5, gemini-3.1 = gemini-3.1-pro, gpt-5.6 = gpt-5.6-sol, qwen-3.8 = qwen-3.8-max, deepseek = deepseek-v4-pro. Field gap +17.0pp on field mean stated confidence 62.6; by month the field gap read +17.8pp / +16.3pp / +17.0pp (Jul 15-31 / Aug 1-31 / Sep 1-30). 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

Every MarketMania forecast carries a model-stated confidence from 0 to 100 alongside its direction call. This report checks whether that number tracks reality: on a well-calibrated forecaster, calls made at 70% confidence should hit their direction about 70% of the time. The monthly series carries one point per model per issue. A calendar quarter carries three monthly points plus a pooled quarterly figure (this first issue has no prior-quarter baseline), which is what the slope table and the month-by-month confidence buckets below are computed from. Descriptive for this window.

How to read this

Gap pp = mean stated confidence minus hit-rate, in percentage points; positive means overconfident. Brier is the mean squared error of the stated probability against the realised outcome, so lower is better and 0.25 is what an uninformative always-50% predictor scores. Coverage = calls scored divided by the mature calls available to that model. Confidence buckets are the models' own natural breakpoints, not an arbitrary binning; cells below N=10 are marked insufficient and never used to rank anything. Prediction and trading metrics live in separate sections and are never combined. 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)). The quarterly rows are one pull at the quarterly cutoff (Thu Oct 1, 2026, 16:00 UTC); the monthly rows are the frozen monthly packs at their own cutoffs, so the two do not have to agree to the last decimal.

Calibration by model (sorted by Brier, best first)

Model	N	Cov.	Hit rate	Mean conf	Gap pp	Brier	95% CI
grok-4.6 (spliced)	7,244	45.9%	45.0%	60.0	+15.0	0.2716	43.9%-46.2%
claude-opus-5 (spliced)	8,261	46.3%	45.9%	60.9	+15.0	0.2720	44.8%-46.9%
claude-fable-5	8,447	47.3%	45.5%	60.7	+15.1	0.2722	44.5%-46.6%
deepseek-v4-pro	7,799	43.7%	45.3%	60.9	+15.6	0.2754	44.2%-46.4%
qwen-3.8-max (spliced)	9,537	53.7%	45.4%	61.5	+16.1	0.2784	44.4%-46.4%
gemini-3.1-pro	9,130	51.4%	47.2%	66.0	+18.9	0.2874	46.1%-48.2%
gpt-5.6-sol	9,188	51.4%	44.9%	67.4	+22.5	0.3014	43.9%-45.9%
Field (all models)	59,606	-	45.6%	62.6	+17.0	0.2803	-

N = directional calls scored this window. Cov. = coverage, scored / mature calls available to that model. Gap pp = mean stated confidence minus hit-rate (positive = overconfident). 95% CI is the descriptive Wilson interval on hit-rate. There are no prior gaps to

compare against and the page-2 chart has no gray series because this is the first quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026). 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).

Calibration curve: hit-rate by stated-confidence bucket

Model	50-60	60-70	70-80	80-100
claude-fable-5	45.6% (2,742)	45.5% (5,703)	50.0% (2) *	n/a (0)
claude-opus-5 (spliced)	47.1% (2,139)	45.5% (6,108)	16.7% (12)	n/a (0)
deepseek-v4-pro	46.0% (2,552)	45.0% (4,743)	43.5% (432)	53.3% (15)
gemini-3.1-pro	35.4% (65)	47.2% (6,279)	47.0% (2,727)	56.9% (58)
gpt-5.6-sol	43.4% (106)	45.9% (6,537)	42.3% (2,530)	40.0% (15)
grok-4.6 (spliced)	45.8% (3,561)	44.3% (3,650)	42.3% (26)	n/a (0)
qwen-3.8-max (spliced)	46.6% (3,077)	44.9% (5,120)	44.3% (1,006)	50.0% (46)

n in parentheses. * = N<10 (insufficient) -- no conclusions drawn from these cells. "n/a (0)" = no calls landed in that bucket this quarter. Sub-50 bucket (n greater than 0 only): grok-4.6 (spliced) 28.6% (n=7) *; claude-opus-5 (spliced) 50.0% (n=2) *; deepseek-v4-pro 54.4% (n=57); qwen-3.8-max (spliced) 43.4% (n=288); gemini-3.1-pro 100.0% (n=1) *. Mid-scale ranking: 2 of the 7 models with N>=10 in both 50-60 and 60-70 out-hit 50-60 from 60-70 this quarter.

Hit-rate by forecast horizon (FH)

Model	1h	4h	1d
claude-fable-5	46.8% (5,347)	44.0% (2,603)	39.8% (497)
claude-opus-5 (spliced)	47.0% (5,085)	44.8% (2,666)	39.6% (510)
deepseek-v4-pro	46.4% (4,875)	44.5% (2,482)	37.8% (442)
gemini-3.1-pro	48.1% (5,832)	46.0% (2,774)	42.9% (524)
gpt-5.6-sol	45.8% (5,744)	44.2% (2,931)	38.4% (513)
grok-4.6 (spliced)	46.2% (4,579)	44.0% (2,292)	35.9% (373)
qwen-3.8-max (spliced)	46.2% (6,018)	44.7% (2,978)	39.2% (541)

n in parentheses; * = N<10 (insufficient). 0 of 7 model lines hit better on 1d than on 1h this quarter.

Trading result (a different question)

A different question: is any of this profitable to trade. "Accurate" and "profitable" are not the same thing, and per methodology this section is never mixed into the prediction/calibration metrics above.

Model	Trades	WR	Net PnL	Gross PnL	Max DD
gemini-3.1-pro	9,130	36.3%	-\$835.51	+\$77.49	-\$835.70
claude-fable-5	8,447	35.6%	-\$888.90	-\$44.20	-\$888.90
deepseek-v4-pro	7,799	36.5%	-\$910.72	-\$130.82	-\$913.30
claude-opus-5 (spliced)	8,261	35.4%	-\$911.08	-\$84.98	-\$911.08
grok-4.6 (spliced)	7,244	34.1%	-\$929.15	-\$204.75	-\$929.15
qwen-3.8-max (spliced)	9,537	35.0%	-\$1087.02	-\$133.32	-\$1087.02
gpt-5.6-sol	9,188	34.8%	-\$1141.58	-\$222.78	-\$1141.65

0 of 7 model lines finished net-positive and 1 gross-positive over the quarter; by month 0 / 0 / 0 lines finished net-positive and the field net PnL ran -\$3095.02 / -\$1054.79 / -\$2554.15 (Jul 15-31 / Aug 1-31 / Sep 1-30). Ranked best-to-worst net PnL. Per methodology this table is never combined with the calibration tables above. The pack labels one of the three months calm and two volatile. The prior-quarter comparison reads n/a because this is the first quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026).

Platform note: TP/SL calibration

Context, not a finding. Since Aug 13 the public sandbox can trade with model-specific TP/SL multipliers learned from this same history (v1); since Aug 19, v2 adds per-ticker and confidence-bucket (50-70 / 70-100) resolution. Both switches landed INSIDE this window, so the quarter spans three feature states. That feature consumes calibration history; it does not feed back into any table in this report, which measures stated confidence vs. direction-hit only. No effect size is claimed for v1 or v2. The toggle lives at marketmania.ai/indices.

Overconfidence drift, month by month

The three months of the quarter are the windows of the published Monthly #1 (its prior window and its month) and Monthly #2, and their monthly gaps read +17.8pp -> +16.3pp -> +17.0pp. Slope is an ordinary least-squares fit over months 1..3; the pack calls a line "flat" when |slope| is under 0.5 pp/month. 3 of 7 lines read narrowing, 2 widening and 2 flat, and the field line reads flat at -0.40 pp/month.

Model	M1 gap	M2 gap	M3 gap	Quarter gap	Slope pp/mo	Trend
claude-opus-5 (spliced)	+15.6	+13.2	+15.9	+15.0	+0.15	flat
grok-4.6 (spliced)	+14.2	+15.1	+15.8	+15.0	+0.80	widening
claude-fable-5	+15.9	+13.9	+15.6	+15.1	-0.15	flat
deepseek-v4-pro	+16.6	+15.4	+15.3	+15.6	-0.65	narrowing
qwen-3.8-max (spliced)	+18.7	+14.7	+14.7	+16.1	-2.00	narrowing
gemini-3.1-pro	+20.0	+18.8	+18.0	+18.9	-1.00	narrowing
gpt-5.6-sol	+22.1	+21.8	+23.5	+22.5	+0.70	widening
Field (all models)	+17.8	+16.3	+17.0	+17.0	-0.40	flat

Months are M1 Jul 15-31, M2 Aug 1-31, M3 Sep 1-30. Slope = OLS over months 1..3, in pp per month; "flat" means |slope| < 0.5 pp/month. Source: the frozen monthly packs (7-line roster, lineage spliced), i.e. the windows of the published Monthly #1 (its prior window and its month) and Monthly #2 at their own cutoffs. The Quarter gap is the pooled quarter, not the mean of the three months; the prior-quarter field gap reads n/a (first quarterly issue). 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).

Bucket	M1	M2	M3
50-60	46.5% vs 56.8 (4,374)	46.1% vs 57.0 (4,952)	45.7% vs 56.9 (4,916)
60-70	44.5% vs 63.3 (12,588)	46.9% vs 63.0 (12,120)	45.5% vs 62.9 (13,432)
70-80	44.4% vs 73.0 (2,262)	43.6% vs 72.9 (2,213)	45.5% vs 72.9 (2,260)
80-100	58.6% vs 81.8 (58)	50.0% vs 81.0 (42)	44.1% vs 80.5 (34)

Field level: every model's calls pooled into the stated-confidence bucket. Cells are hit rate vs mean stated confidence, with n in parentheses; * = N<10 (insufficient), "n/a (0)" = no call landed in that bucket that month. Source: the quarter pull split by slot month at the quarterly cutoff, so the three columns are one consistent pool.

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

- Pooled over the quarter the field's gap is +17.0pp on 59,606 scored calls: stated confidence averaged 62.6 while the direction hit 45.6%.
- The field gap spans 1.5 pp across the three months inside the quarter, against a pooled quarterly gap of +17.0pp on field mean stated confidence of 62.6 (by month 62.8 / 62.6 / 62.6). The prior-quarter gap reads n/a because this is the first quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026).
- Trading over the quarter finished 0 of 7 lines net-positive (by month 0 / 0 / 0); the pack labels one of the three months calm and two volatile. The prior-quarter comparison reads n/a because this is the first quarterly issue: there is no prior quarter (platform data begin Jul 15, 2026).

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.
- Prediction metrics (this report) and trading metrics are kept in separate sections per methodology -- they are never combined into a single score. 95% Wilson CIs shown are descriptive, not inferential: observations inside one window are dependent, so read them as a range, not a formal coverage guarantee.

- All 59,606 scored calls sit inside one calendar quarter, and three monthly points are three points: an OLS slope over them is a description of this quarter, not evidence of drift. 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. Models report confidence at discrete levels, not a continuous scale; the buckets reflect those natural breakpoints and cells below N=10 are marked insufficient.
- Underlying price series are reconstructed from trade entry prices (median per symbol-slot), not an independent tick feed. 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: 59,606 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 · 48 published reports.

Issue #1. Quarterly Calibration is a living quarterly comparison; each issue appends one more calendar quarter of confidence-vs-hit data. This first issue sets the baseline on a partial quarter (Jul 15-Sep 30): field gap +17.0pp on 59,606 scored calls, field Brier 0.2803, monthly gaps +17.8pp -> +16.3pp -> +17.0pp, field slope -0.40 pp/month (flat); 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 Consensus Watch #1	https://marketmania.ai/research/reports/consensus-watch-quarterly-2026-q3.pdf
Quarterly Model Watch #1	https://marketmania.ai/research/reports/model-watch-quarterly-2026-q3.pdf
Monthly Calibration #2	https://marketmania.ai/research/reports/calibration-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_calibration_quarterly_2026q3,
  title = {Quarterly Calibration #1: confidence vs. direction-hit, Jul 15-Sep 30 2026},
  author = {{MarketMania Research}},
  year  = {2026}, month = {October}, day = {8},
  url   = {https://marketmania.ai/research/reports/calibration-quarterly-2026-q3.pdf},
  note  = {Methodology v1.1, hash e66c7e8c864a2233; source quarterly_metrics_2026-q3.json}
}
```