Monthly Calibration #3

November 5, 2026 · MarketMania Research · Monthly series · window Oct 1-31, 2026 UTC

RESEARCH SNAPSHOT			
22,579 directional calls scored	**7** models (lineage spliced: opus, gpt, grok)	of **40,889** mature forecasts	window **Oct 1-31, 2026 UTC**
hit = direction rule: tp1/tp2->hit, sl->miss, expiry->sign of gross pnl	FH gate **1h / 4h / 1d**	base field hit **48.6%** vs **61.8** stated	cutoff **Sun Nov 1, 2026, 16:00 UTC**
market: BTC net **-7.22%** (prior +6.42%)	ann vol **28.3%** (was 41.2%)	TOP5 volume **$68.94B**, -14% m/m	pairwise corr **0.87** (was 0.78)

KEY FINDING · OBSERVATION (one monthly window)

Over the month the field's overconfidence gap read +13.2pp against 61.8 stated mean confidence, the weekly gaps ran +4.5pp to +19.6pp inside it, and the field slope across the four weeks is -0.67 pp/week (narrowing).

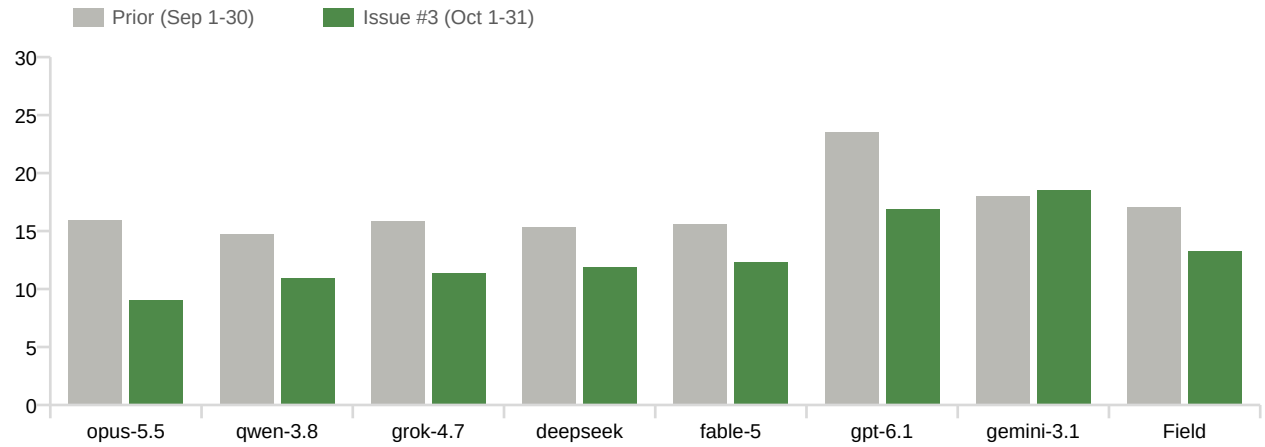
WEEKLY EVIDENCE -> MONTHLY VERDICT (rule below)

Metric	W1 Oct 5-11	W2 Oct 12-18	W3 Oct 19-25	W4 Oct 26-Nov 1	Month	Verdict
Field gap (conf - hit), pp	+18.2	+15.4	+4.5	+19.6	+13.2	Confirmed (4/4)
Lowest-Brier model	opus-5.5	deepseek	opus-5.5	opus-5.5	opus-5.5	Confirmed (3/4)
Widest-gap model	gpt-6.1	gpt-6.1	gemini-3.1	gemini-3.1	gemini-3.1	Inconclusive (2/4)

Verdict rule: each row states one reading of the month; a week agrees when it shows the same reading (same sign, or the same name). Confirmed = 3 or 4 of the 4 weeks agree with the month; Inconclusive = 2; Rejected = 0 or 1. Weekly cells come from the frozen weekly packs (7-line roster); the month is the pooled Oct 1-31 pull. W4 straddles the month edge (Nov 1); Oct 1-4 belong to the month only.
The model cutover of Oct 17, 2026 09:01 UTC (gpt-5.6-sol -> gpt-6.1-sol, claude-opus-5 -> claude-opus-5.5, grok-4.6 -> grok-4.7) falls inside W2; W1 runs on the earlier builds, W3 and W4 on the new ones, the month column joins both -- all per-line rows are model lines, not single builds.

TL;DR

- **OBSERVATION -- The monthly gap read +13.2pp.** Directional calls hit 48.6% against 61.8 stated mean confidence over 22,579 calls. Field Brier is 0.2701 and 0 of 7 model lines scored below 0.25, the score of an uninformative always-50% predictor. One month, descriptive.
- **The field gap slope across the four weeks is -0.67 pp/week (narrowing).** Across those weeks the field gap ran +18.2pp -> +15.4pp -> +4.5pp -> +19.6pp and the per-model slopes split 4 narrowing / 1 widening / 2 flat under the pack's own rule (|slope| < 0.5 pp/week = flat). claude-opus-5.5 (spliced) is both the lowest-Brier line (Brier 0.2600) and the hit-rate leader (49.6%).
- **The widest weekly swing belongs to gpt-6.1-sol (spliced).** Its weekly gap moved over a 19.1 pp range inside one month, against a field range of 15.1 pp. Gap moves against the prior window ran from -6.9pp for claude-opus-5.5 (spliced) to +0.5pp for gemini-3.1-pro; 6 of 7 lines sit below their prior-window gap.
- **Trading, kept separate from every calibration table.** 0 of 7 lines finished net-positive over the month; net PnL ran -$161.57 (grok-4.7 (spliced), best) to -$351.44 (gemini-3.1-pro, worst), field -$1648.92.



Overconfidence gap in percentage points (mean stated confidence minus hit-rate), the prior window against the whole month, sorted by this month's gap. Short names: opus-5.5 = claude-opus-5.5, fable-5 = claude-fable-5, gemini-3.1 = gemini-3.1-pro, gpt-6.1 = gpt-6.1-sol, qwen-3.8 = qwen-3.8-max, deepseek = deepseek-v4-pro. Field gap +17.0pp -> +13.2pp; field mean stated confidence 62.6 -> 61.8. The prior window is the full September: window A of the same pull, slots 2026-09-01 to 2026-10-01 (first slot actually seen 2026-09-01 00:01 UTC), matured at the prior cutoff Thu Oct 1, 2026, 16:00 UTC under the same engine, the same FH gate and the same hit rule as October -- 20,682 directional calls against 22,579 in the month.

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 weekly series carries one point per model per issue. A calendar month carries four weekly points plus a pooled monthly figure and a prior-window baseline, which is what the slope table and the week-by-week 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 month (claude-opus-5.5 (incl. opus-5-era, Oct 1-17); gpt-6.1-sol (incl. 5.6-era, Oct 1-17); grok-4.7 (incl. 4.6-era, Oct 1-17)). The monthly rows are one pull at the monthly cutoff (Sun Nov 1, 2026, 16:00 UTC); the weekly rows are the published issues at their own frozen 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
claude-opus-5.5 (spliced)	2,612	44.7%	49.6%	58.6	+9.0	0.2600	47.7%-51.5%
grok-4.7 (spliced)	3,069	53.5%	48.6%	60.0	+11.3	0.2631	46.9%-50.4%
qwen-3.8-max	3,344	56.9%	48.4%	59.3	+10.9	0.2631	46.7%-50.0%
deepseek-v4-pro	3,485	59.5%	48.5%	60.4	+11.9	0.2653	46.9%-50.2%
claude-fable-5	3,003	51.2%	48.2%	60.5	+12.3	0.2661	46.4%-50.0%
gpt-6.1-sol (spliced)	3,627	62.0%	48.9%	65.8	+16.9	0.2816	47.3%-50.5%
gemini-3.1-pro	3,439	58.7%	48.4%	66.9	+18.5	0.2872	46.7%-50.1%
Field (all models)	**22,579**	-	**48.6%**	**61.8**	**+13.2**	**0.2701**	-

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. Prior gaps per model are the gray bars on page 1: prior window Sep 1-30, field gap +17.0pp against +13.2pp here, on n=20,682 directional calls there.

claude-opus-5.5 (spliced) = claude-opus-5.5 (incl. opus-5-era, Oct 1-17). gpt-6.1-sol (spliced) = gpt-6.1-sol (incl. 5.6-era, Oct 1-17). grok-4.7 (spliced) = grok-4.7 (incl. 4.6-era, Oct 1-17). In the prior window the lineage splice folds 6,300 claude-opus-5 rows into claude-opus-5.5, 6,300 gpt-5.6-sol rows into gpt-6.1-sol and 6,300 grok-4.6 rows into grok-4.7 (18,900 rows in all).

Calibration curve: hit-rate by stated-confidence bucket

Model	50-60	60-70	70-80	80-100
claude-fable-5	47.9% (1,008)	48.4% (1,993)	0.0% (2) *	n/a (0)
claude-opus-5.5 (spliced)	51.4% (1,524)	47.0% (1,088)	n/a (0)	n/a (0)
deepseek-v4-pro	48.0% (1,253)	48.8% (2,171)	50.8% (59)	n/a (0)
gemini-3.1-pro	65.0% (20)	47.8% (2,161)	49.6% (1,152)	44.3% (106)
gpt-6.1-sol (spliced)	49.5% (317)	48.9% (2,542)	48.3% (753)	60.0% (15)
grok-4.7 (spliced)	47.6% (1,254)	49.5% (1,809)	0.0% (4) *	n/a (0)
qwen-3.8-max	48.0% (1,664)	48.7% (1,619)	57.1% (14)	n/a (0)

n in parentheses. * = N<10 (insufficient) -- no conclusions drawn from these cells. "n/a (0)" = no calls landed in that bucket this month. Sub-50 bucket (n greater than 0 only): grok-4.7 (spliced) 0.0% (n=2) *; qwen-3.8-max 48.9% (n=47); deepseek-v4-pro 50.0% (n=2) *. Mid-scale ranking: 4 of the 7 models with N>=10 in both 50-60 and 60-70 out-hit 50-60 from 60-70 this month.

Hit-rate by forecast horizon (FH)

Model	1h	4h	1d
claude-fable-5	49.2% (1,906)	46.6% (934)	45.4% (163)
claude-opus-5.5 (spliced)	50.6% (1,616)	47.4% (852)	51.4% (144)
deepseek-v4-pro	50.1% (2,103)	45.9% (1,185)	48.2% (197)
gemini-3.1-pro	50.3% (2,179)	45.0% (1,070)	45.8% (190)
gpt-6.1-sol (spliced)	49.8% (2,189)	47.2% (1,213)	49.8% (225)
grok-4.7 (spliced)	49.1% (1,867)	46.5% (1,028)	56.3% (174)
qwen-3.8-max	48.9% (2,082)	46.6% (1,085)	53.1% (177)

n in parentheses; * = N<10 (insufficient). 4 of 7 model lines hit better on 1d than on 1h this month.

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
grok-4.7 (spliced)	3,069	40.3%	-$161.57	+$145.33	-$213.21
claude-opus-5.5 (spliced)	2,612	40.5%	-$163.90	+$97.30	-$184.65
qwen-3.8-max	3,344	39.5%	-$183.77	+$150.63	-$214.03
deepseek-v4-pro	3,485	40.9%	-$236.63	+$111.87	-$260.48
claude-fable-5	3,003	38.8%	-$271.41	+$28.89	-$282.31
gpt-6.1-sol (spliced)	3,627	39.5%	-$280.19	+$82.51	-$288.45
gemini-3.1-pro	3,439	38.1%	-$351.44	-$7.54	-$363.23

0 of 7 model lines finished net-positive and 6 gross-positive over the month, against 0 of 7 net-positive in the prior window Sep 1-30 on a summed net PnL of -$2554.16. Ranked best-to-worst net PnL. Per methodology this table is never combined with the calibration tables above. The pack labels two of the four weeks calm and two volatile.

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 before this window and before the prior window (Aug 13 and Aug 19). 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, week by week

The four Mon-Sun weeks from Oct 5 to Nov 1 are the published weekly issues #10, #11, #12 and #13, and their weekly gaps read +18.2pp -> +15.4pp -> +4.5pp -> +19.6pp. Slope is an ordinary least-squares fit over weeks 1..4; the pack calls a line "flat" when |slope| is under 0.5 pp/week. 4 of 7 lines read narrowing, 1 widening and 2 flat, and the field line reads narrowing at -0.67 pp/week.

Model	W1 gap	W2 gap	W3 gap	W4 gap	Month gap	Slope pp/wk	Trend
claude-opus-5.5 (spliced)	+13.8	+13.4	-1.5	+15.4	+9.0	-1.01	narrowing
qwen-3.8-max	+15.4	+12.7	+3.5	+17.7	+10.9	-0.23	flat
grok-4.7 (spliced)	+16.3	+12.9	+4.3	+17.1	+11.3	-0.62	narrowing
deepseek-v4-pro	+18.1	+11.6	+4.5	+18.8	+11.9	-0.50	narrowing
claude-fable-5	+14.9	+14.4	+4.5	+20.1	+12.3	+0.57	widening
gpt-6.1-sol (spliced)	+24.7	+21.5	+5.6	+20.7	+16.9	-2.79	narrowing
gemini-3.1-pro	+21.8	+20.0	+9.3	+26.7	+18.5	+0.40	flat
Field (all models)	**+18.2**	**+15.4**	**+4.5**	**+19.6**	**+13.2**	**-0.67**	**narrowing**

Weeks are W1 Oct 5-11, W2 Oct 12-18, W3 Oct 19-25, W4 Oct 26-Nov 1. Slope = OLS over weeks 1..4, in pp per week; "flat" means |slope| < 0.5 pp/week. Source: the frozen weekly packs (7-line roster), i.e. the windows of the published weekly issues #10, #11, #12 and #13 at their own cutoffs. The Month gap is the pooled month, not the mean of the four weeks; the prior window Sep 1-30 read a field gap of +17.0pp. claude-opus-5.5 (spliced) = claude-opus-5.5 (incl. opus-5-era, Oct 1-17). gpt-6.1-sol (spliced) = gpt-6.1-sol (incl. 5.6-era, Oct 1-17). grok-4.7 (spliced) = grok-4.7 (incl. 4.6-era, Oct 1-17). In the prior window the lineage splice folds 6,300 claude-opus-5 rows into claude-opus-5.5, 6,300 gpt-5.6-sol rows into gpt-6.1-sol and 6,300 grok-4.6 rows into grok-4.7 (18,900 rows in all).

Bucket	W1	W2	W3	W4
50-60	44.2% vs 56.9 (1,331)	48.8% vs 56.9 (1,518)	55.3% vs 56.8 (1,967)	43.0% vs 56.8 (1,708)
60-70	43.2% vs 62.7 (2,801)	45.4% vs 62.6 (3,020)	57.7% vs 62.7 (3,450)	41.1% vs 62.7 (2,444)
70-80	45.5% vs 73.1 (437)	46.9% vs 73.1 (527)	59.9% vs 72.5 (404)	40.3% vs 72.7 (320)
80-100	63.6% vs 81.1 (11)	40.0% vs 81.1 (10)	50.0% vs 81.5 (26)	37.7% vs 84.2 (53)

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 week. Source: the month pull split by slot week at the monthly cutoff, so the four columns are one consistent pool. W4 holds Oct 26-31 here: the month pull ends on Oct 31.

Market check: October ran calmer than September

OBSERVATION — mean |1d move| 1.91% -> 1.57% (-17% rel), BTC realized vol 34.9% -> 32.4% (ann., hourly); field directional accuracy 45.5% -> 48.6% (+3.1pp), 6 of 7 models improved, sim win-rate up for 6 of 7, field sim PnL -$2554.15 -> -$1648.92 (Sep 1-30 vs Oct 1-31; descriptive, one pair of windows, not a claim).
Robustness: raw price-sign accuracy 48.4% -> 48.1% — the opposite direction to the trade-based hit rule.

Measure	Sep 1-30	Oct 1-31	Change
Mean |1d move| (5 assets)	1.91%	1.57%	-17% rel
BTC realized vol (ann., hourly)	34.9%	32.4%	-2.5pp
Field directional accuracy	45.5%	48.6%	+3.1pp
Models improving hit-rate	--	6 of 7	--
Raw price-sign accuracy	48.4%	48.1%	-0.3pp
Field sim win-rate	38.4%	39.6%	+1.2pp
Field sim net PnL	-$2554.15	-$1648.92	--

Window A is the pack's prior window -- the full September -- so this slice and every month-over-month column in this report compare October against the full September. Descriptive, one pair of 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). 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-5.5 (incl. opus-5-era, Oct 1-17); gpt-6.1-sol (incl. 5.6-era, Oct 1-17); grok-4.7 (incl. 4.6-era, Oct 1-17). 10,395 rows were remapped inside the month against 18,900 in the prior window.

Practical implications

- Pooled over a whole month the field's gap is +13.2pp on 22,579 scored calls: stated confidence averaged 61.8 while the direction hit 48.6%.
- The field gap spans 15.1 pp across the four weeks set against the month, against a pooled monthly gap of +13.2pp and a prior-window gap of +17.0pp; field mean stated confidence moved 62.6 -> 61.8 over the same pair of windows.
- Trading this month finished 0 of 7 lines net-positive against 0 of 7 in the prior window; the pack labels two of the four weeks calm and two volatile.

Limitations

- The prior window is the full September: every "prior" column, the gray series of the page-1 chart and the market-check A column describe Sep 1-30. It is window A of the same pull, slots 2026-09-01 to 2026-10-01 (first slot actually seen 2026-09-01 00:01 UTC), matured at the prior cutoff Thu Oct 1, 2026, 16:00 UTC under the same engine, the same FH gate and the same hit rule as October -- 20,682 directional calls against 22,579 in the month. Month-over-month deltas are a step between two windows of different length (30 and 31 days).
- Two cutoffs live in this issue. The month is frozen at Sun Nov 1, 2026, 16:00 UTC; the four weekly blocks are frozen at their own published cutoffs (W1 Oct 5-11, W2 Oct 12-18, W3 Oct 19-25, W4 Oct 26-Nov 1). A forecast still pending at its week's cutoff is immature in the weekly pack and mature in the month pull, so the month does not equal the sum of the weeks (22,579 vs 20,635 directional, plus 2,505 on the 4 remainder days); W4 also runs to Nov 1, a day the month pull does not hold, so its pack counts 5,094 directional calls against 4,533 for Oct 26-31 inside the month pull -- that gap is arithmetic, not a data problem.
- 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 22,579 scored calls sit inside one calendar month, and four weekly points are four points: an OLS slope over them is a description of this month, not evidence of drift. This window ran 9 trend days of 31, and the pack's own volatility rule splits the four Mon-Sun weeks from Oct 5 to Nov 1 as W1 volatile, W2 calm, W3 volatile, W4 calm; every month-over-month comparison 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 (monthly #3): (1) rolling-1w forecasts emit daily since Aug 22, 2026 and rolling-1M daily since Aug 23, 2026, so across this 31-day window daily coverage is FULL as measured -- the 1w series has slots on 31 of 31 days and the 1M series on 31 of 31 days; both sit outside this report's FH gate (1h/4h/1d) and appear only in the exclusions counter (1w 2,160, 1M 2,135). (2) Model lineage is a splice inside this window. Three lines flipped at Oct 17, 2026 09:01 UTC, day 17 of 31 of the month: "gpt-6.1-sol (incl. 5.6-era, Oct 1-17)" (3,465 rows from the gpt-5.6-sol era, 3,045 from gpt-6.1-sol); "claude-opus-5.5 (incl. opus-5-era, Oct 1-17)" (3,465 rows from the claude-opus-5 era, 3,045 from claude-opus-5.5); "grok-4.7 (incl. 4.6-era, Oct 1-17)" (3,465 rows from the grok-4.6 era, 3,045 from grok-4.7). Every per-model aggregate sees the merged head id (the raw id is kept in the pack); 10,395 rows were remapped inside the month, and in the prior window 6,300 rows of each of the three lines (18,900 in all) are earlier-build rows, so each month-over-month row for those lines joins a spliced month to a single-build prior month -- a remapped comparison is not the same model build. (3) The two hourly slots before the flip (Oct 17, 07:01 and 08:01 UTC) hold 60 invalid rows (claude-opus-5 20, gpt-5.6-sol 20, grok-4.6 20), counted in the invalid counter as they are; in those two hours the other 4 lines (claude-fable-5, deepseek-v4-pro, gemini-3.1-pro, qwen-3.8-max) carry ok rows. (4) Past lineage flips: the grok 4.5 -> 4.6 flip (Aug 24, 2026 09:22 UTC) and the qwen-3.7-max -> qwen-3.8-max succession (Aug 5, 2026) both sit before the prior window.
- 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 prior-month block is the full September: it covers 30 days (Sep 1-30), and the TOP5 volume figure compares a 31-day sum with a 30-day sum -- a level difference, not a like-for-like change.
- Research-to-date counter: directional forecasts resolved inside the FH gate (1h/4h/1d) since Jul 11, counted at the issue cutoff (Sun Nov 1, 2026, 16:00 UTC) -- the D2 definition pinned in the weekly series; this issue prints 82,334. The pack carries no prior-cutoff pin for a monthly window (the prior-cutoff pin is null by construction), so the control this issue is the weekly reproduction: the W4 block of this pack (Oct 26-Nov 1) reproduces the published weekly #13 on 7 of 7 checks, and the prior window of this pack (Sep 1-30, prior cutoff Thu Oct 1, 2026, 16:00 UTC) reproduces the published Monthly #2 on 7 of 7 checks. Counter deltas against pre-wave-3 issues remain definitional.

Counters & lineage

Counter	Value	Counter	Value
Forecasts total	45,540	Mature (scored pool)	40,889
OK in gate	40,889	-- of them directional	22,579
Out of gate (1w / 1M)	2,160 / 2,135	-- of them sideways	18,310
Invalid	356	Pending (next issue)	0
Late closes	0	Uptime, all grids	1,178 / 1,178 slots
Source file	monthly_metrics_2026-10.json	Report cutoff	Sun Nov 1, 2026, 16:00 UTC
Generated at (pipeline)	2026-11-03 12:39 UTC	Methodology	v1.1 (2026-08-10) · hash e66c7e8c864a2233

THIS REPORT: 22,579 scored observations. MARKETMANIA RESEARCH TO DATE (as of cutoff Nov 1, 16:00 UTC): 82,334 resolved since Jul 11 · 7 models tracked (current line-up; earlier versions folded into their successors' lineage) · 5 assets · 5 horizons · hourly · 68 published reports.

Issue #3. Monthly Calibration is a living monthly comparison; each issue appends one more calendar month of confidence-vs-hit data. This third issue reads October against the full September: field gap +13.2pp on 22,579 scored calls, field Brier 0.2701, weekly gaps +18.2pp -> +15.4pp -> +4.5pp -> +19.6pp, field slope -0.67 pp/week (narrowing), against a prior-window field gap of +17.0pp. 3 model lines are lineage splices this month (claude-opus-5.5 (incl. opus-5-era, Oct 1-17); gpt-6.1-sol (incl. 5.6-era, Oct 1-17); grok-4.7 (incl. 4.6-era, Oct 1-17)). Engine 1.1 has powered the sandbox since Aug 18, i.e. before this window and before the prior window; no cross-engine PnL comparisons are claimed.

What we're testing next

- Monthly #4 (November): the same three series over Nov 1-30, cut on Dec 1 plus the maturity lag, so its month-over-month comparison runs against this October -- the first month whose prior window carries the cutover splice.
- Monthly Config Watch #3: the same October window through the config search engine, published as the fourth monthly report of this issue.
- Quarterly Stability Index: the next repeat-prompt stability run is the next cross-quarter point on model stability; its date is announced with the run.

Related research

Report	Direct PDF link
Monthly Consensus Watch #3	https://marketmania.ai/research/reports/consensus-watch-monthly-2026-10.pdf
Monthly Model Watch #3	https://marketmania.ai/research/reports/model-watch-monthly-2026-10.pdf
Weekly Calibration #13	https://marketmania.ai/research/reports/weekly-calibration-2026-10-26.pdf

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

CITE THIS REPORT

```
@misc{mm_calibration_monthly_2026m10,
  title  = {Monthly Calibration #3: confidence vs. direction-hit, Oct 1-31 2026},
  author = {{MarketMania Research}},
  year   = {2026}, month = {November}, day = {5},
  url    = {https://marketmania.ai/research/reports/calibration-monthly-2026-10.pdf},
  note   = {Methodology v1.1, hash e66c7e8c864a2233; source monthly_metrics_2026-10.json}
}
```