

Monthly Model Watch #1

September 4, 2026 · MarketMania Research · Monthly series · window Aug 1-31, 2026 UTC

RESEARCH SNAPSHOT			
19,383 directional calls scored	7 models (lineage spliced: grok, qwen)	of 40,979 mature forecasts	window Aug 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 46.3%	cutoff Mon Sep 1, 2026, 16:00 UTC
market: BTC net +24.95% (prior -3.31%)	ann vol 41.5% (was 27.3%)	TOP5 volume \$65.97B, +136% m/m	pairwise corr 0.79 (was 0.79)

KEY FINDING · OBSERVATION (one monthly window)

Field accuracy went 45.0% -> 46.3% (+1.4pp) from the partial July to August with 5 of 7 models improving, 3 distinct model lines led the four weeks, and no monthly title was awarded.

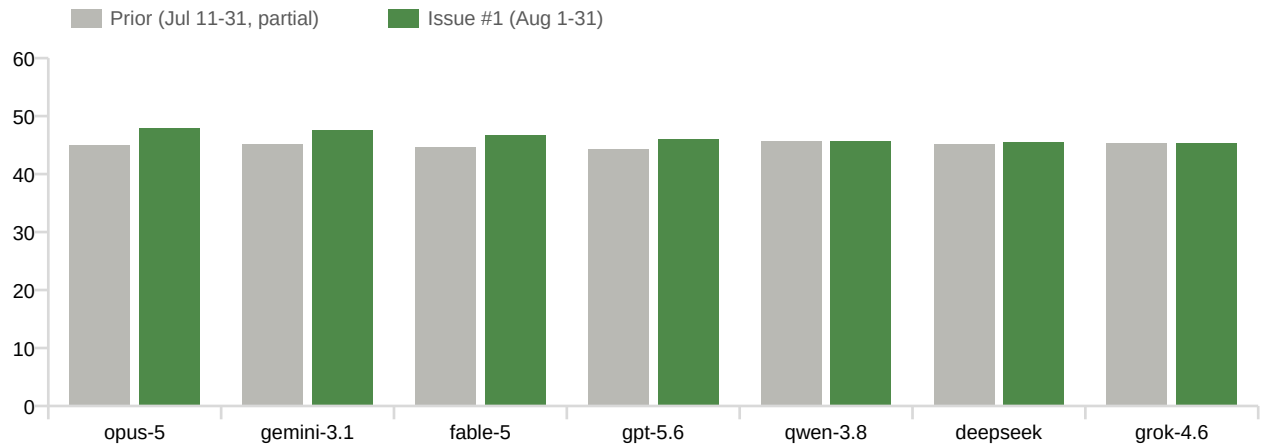
WEEKLY EVIDENCE -> MONTHLY VERDICT (rule below)

Metric	W1 Aug 3-9	W2 Aug 10-16	W3 Aug 17-23	W4 Aug 24-30	Month	Verdict
Leader (hit-rate)	gemini-3.1	gemini-3.1	opus-5	deepseek	opus-5	Rejected (1/4)
Most readable coin	SOL	BNB	BTC	SOL	BNB	Rejected (1/4)
Field hit below 50%	42.4%	43.8%	54.0%	42.9%	46.3%	Confirmed (3/4)

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, lineage spliced); the month is the pooled Aug 1-31 pull.

TL;DR

- **OBSERVATION -- No monthly title.** claude-opus-5 leads the month descriptively at 47.8% [45.7%, 49.8%]; the gap to #2 (gemini-3.1-pro, 47.6%) is 0.2pp. 5 of 7 models improved against the partial July, and 3 distinct model lines held the weekly top spot across the four weeks.
- **No ticker cleared 51%.** BNB is the month's most readable coin (48.1%, n=3,857) and ETH the hardest at 42.9%; 0 of the 5 assets cleared 51% over the month. Pair of the month: claude-opus-5 x SOL 50.4% (n=482); worst: qwen-3.8-max (spliced) x ETH 40.5% (n=576).
- **Mean weekly rank ran 1.75 to 5.50 across the four weeks.** claude-opus-5 swung across 6 rank positions and gemini-3.1-pro across 2. Consecutive-week Spearman came in at W1-W2 0.96, W2-W3 -0.32, W3-W4 0.07.
- **The field read volatile weeks better.** Split by the pack's own volatility rule the field hit 43.1% (n=7,712) in the two calm weeks against 48.8% (n=9,706) in the two volatile ones, a difference of +5.7pp. Cross-TF/FH lifts ran -5.6pp to -0.6pp. On trend days, 4h TF-agreement hit 56.6% (n=749) against 39.0% (n=1,240) on flat days.



Directional hit-rate (%) by model, the prior window against the whole month, ranked by this month's rate. 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 hit 45.0% -> 46.3%. The prior window is PARTIAL: window A of the same pull, slots 2026-07-01 to 2026-08-01, forecasts counted from Jul 11 (first slot actually seen 2026-07-15 18:01 UTC) and audited candles from Jul 15, matured at the prior cutoff Sat Aug 1, 2026, 16:00 UTC under the same engine, the same FH gate and the same hit rule as August -- 19,541 directional calls against 19,383 in the month. The leaderboard's M/m column is computed on unrounded rates and can differ by 0.1pp from the difference of the rounded columns.

Why it matters

MarketMania scores 7 model lines against the same market, hour after hour. Monthly Model Watch asks the weekly questions at monthly n -- who led, which tickers were readable, who saw the turn first, does self-agreement help -- and then asks the three questions a single week cannot answer: which coin the field reads best over a month, how far a weekly rank moves inside one month, and how the field's accuracy differs between the calm and volatile weeks the pack labels. Each block gets its own table, its own N and its own caveats.

How to read this

Every block below is scored on the same pool: matured directional calls inside the 1h/4h/1d FH gate, hit by the methodology v1.1 direction rule, over the whole month at one cutoff (Mon Sep 1, 2026, 16:00 UTC). The monthly title is awarded only on a >=5pp gap AND non-overlapping 95% Wilson CIs with N>=10 per side -- a descriptive lead is not a title. Reversals use the 4h grid with a >=2.0% counter-move sustained 8h, BTC/ETH only in v1. Cross-TF/FH blocks ask whether a model agreeing with itself does any better, and are split trend vs flat. The by-week blocks come from the four frozen weekly packs at their own published cutoffs. 2 model lines are lineage splices this month (grok-4.6 (incl. 4.5-era, Aug 1-24); qwen-3.8-max (incl. 3.7-era, Aug 1-5)).

Leaderboard of the month

Model	N	Cov.	Hit rate	95% CI	M/m pp
claude-opus-5	2,281	38.9%	47.8%	45.7%-49.8%	+2.9
gemini-3.1-pro	3,033	51.8%	47.6%	45.8%-49.4%	+2.5
claude-fable-5	2,812	48.0%	46.7%	44.9%-48.6%	+2.2
gpt-5.6-sol	2,970	50.6%	45.9%	44.1%-47.6%	+1.6
qwen-3.8-max (spliced)	3,001	51.7%	45.6%	43.9%-47.4%	+0.0
deepseek-v4-pro	2,578	44.0%	45.5%	43.6%-47.5%	+0.5
grok-4.6 (spliced)	2,708	46.2%	45.3%	43.4%-47.1%	+0.0

No monthly title is awarded this issue. Title rule: >=5pp gap AND non-overlapping 95% CI, N>=10 per side. claude-opus-5's lead over gemini-3.1-pro is 0.2pp and their CIs overlap, so neither condition is met. M/m pp = this month's hit rate minus the same line's hit rate in the pack's prior window Jul 11-31 (partial), n=19,541 directional calls there against 19,383 here, so it is a step between windows of different length. grok-4.6 (spliced) = grok-4.6 (incl. 4.5-era, Aug 1-24). qwen-3.8-max (spliced) = qwen-3.8-max (incl. 3.7-era, Aug 1-5).

Ticker of the month

Symbol	N	Hit rate	95% CI	Prior
BNB	3,857	48.1%	46.5%-49.6%	45.5%
SOL	3,928	47.3%	45.8%-48.9%	43.2%
XRP	3,835	46.7%	45.1%-48.3%	44.5%
BTC	4,068	46.5%	45.0%-48.0%	44.2%
ETH	3,695	42.9%	41.3%-44.5%	47.8%

BNB was the field's most readable ticker (48.1%) and ETH the hardest (42.9%); 0 of 5 assets cleared 51% this month. Prior = the same per-symbol hit rate over Jul 11-31 (partial), on n=2,720 to 4,564 per coin. The prior window carried 10 assets; only the current five are compared. This issue's ordering is BNB > SOL > XRP > BTC > ETH.

Pair of the month

Top 6 of the 12 model x ticker pairs tracked this month (best hit-rate), then the 5 worst in their own table -- the pair list is split rather than paged so neither table breaks across a page. Pairs shown require N of 8 or more calls.

Model (top 6)	Symbol	N	Hit rate
claude-opus-5	SOL	482	50.4%
gemini-3.1-pro	XRP	640	49.7%
gemini-3.1-pro	BNB	568	49.6%
qwen-3.8-max (spliced)	BNB	587	49.6%
claude-opus-5	BNB	447	48.5%
claude-fable-5	SOL	604	48.5%

Model (5 worst)	Symbol	N	Hit rate
claude-fable-5	ETH	529	44.0%
grok-4.6 (spliced)	ETH	512	42.0%
gpt-5.6-sol	ETH	574	41.6%
deepseek-v4-pro	ETH	522	41.2%
qwen-3.8-max (spliced)	ETH	576	40.5%

claude-opus-5 x SOL (50.4%, n=482) is the month's best pair and qwen-3.8-max (spliced) x ETH (40.5%, n=576) the weakest. Pair history accumulates across issues -- read this as the first monthly data point, not a ranking.

Who saw the reversal first

Reversal definition: 4h grid; trend = sign of prior 24h; counter-move of 2.0% or more (BTC/ETH) sustained 8h. A model 'calls' the reversal if it has a matured directional hit call in the new direction, FH 4h or 1d, in the slot window [T-12h, T+FH].

Symbol	Time (UTC)	New side	Move (8h)	Callers	First caller
BTC	Mon Aug 3, 08:00	Long	+2.10%	9	qwen-3.8-max (spliced)
BTC	Fri Aug 28, 12:00	Short	-2.49%	8	gpt-5.6-sol
ETH	Sat Aug 1, 20:00	Long	+2.27%	0	NOBODY
ETH	Mon Aug 10, 08:00	Short	-2.65%	0	NOBODY
ETH	Sat Aug 22, 00:00	Short	-3.46%	1	gemini-3.1-pro
ETH	Sun Aug 23, 08:00	Long	+2.14%	17	claude-fable-5
ETH	Wed Aug 26, 16:00	Long	+2.31%	13	gpt-5.6-sol
ETH	Fri Aug 28, 12:00	Short	-2.80%	0	NOBODY
ETH	Sun Aug 30, 16:00	Short	-2.52%	1	deepseek-v4-pro

The BTC long turn drew 9 callers, first qwen-3.8-max (spliced) on a 4h/1h call at Aug 2, 20:01 with confidence 68; the BTC short turn drew 8 callers, first gpt-5.6-sol on a 4h/1h call at Aug 28, 04:01 with confidence 62; the ETH long turn drew NOBODY; the ETH

short turn drew NOBODY; the ETH short turn drew 1 caller, first gemini-3.1-pro on a 4h/1h call at Aug 22, 00:01 with confidence 60; the ETH long turn drew 17 callers, first claude-fable-5 on a 1d/1d call at Aug 23, 00:01 with confidence 62; the ETH long turn drew 13 callers, first gpt-5.6-sol on a 1d/4h call at Aug 27, 00:01 with confidence 69; the ETH short turn drew NOBODY; the ETH short turn drew 1 caller, first deepseek-v4-pro on a 4h/4h call at Aug 30, 04:01 with confidence 56. Scope in v1 remains BTC/ETH only; 9 events in this window, against 1 / 1 / 2 / 4 in the four published weekly issues (W1 Aug 3-9, W2 Aug 10-16, W3 Aug 17-23, W4 Aug 24-30), and 6 of the 9 found at least one caller.

Cross-TF and cross-FH confirmation: does agreeing with yourself help?

Block	FH	vs.	Agree n	Agree hit	Dis n	Dis hit	Base hit	Lift
cross-TF	4h	TF 4h vs 1h	1,989	45.6%	91	34.1%	46.2% (3,086)	-0.6
cross-TF	1d	TF 1d vs 4h	380	50.8%	23	87.0%	54.0% (539)	-3.2
cross-FH	4h	by FH 1h	1,824	45.3%	122	40.2%	46.2% (3,086)	-0.8
cross-FH	1d	by FH 4h	316	48.4%	32	84.4%	54.0% (539)	-5.6

* = N<10 (insufficient); insufficient cells are never used to rank models. Mixed pairs (a model with only one side of the comparison present): 993 / 136 / 1137 / 191. Pooled across all four cell families the lifts run -5.6pp to -0.6pp this month.

Regime control. Methodology calls for a trend-vs-flat split on this table. This window has 9 trend days of 31 (a day is a trend day when its absolute BTC 1d move is 1.5% or more, else a flat day), and the four weeks label as W1 calm, W2 calm, W3 volatile, W4 volatile under the pack's rule (a week is volatile when its mean absolute BTC 1d move is at or above the median of the four weeks, else calm; median mean |BTC 1d move| 0.867%). On trend days, 4h TF-agreement hit 56.6% (n=749) against 39.0% (n=1,240) on flat days.

Side-mix and the month's failure

Model	N	Long	Short	Sideways	L:S
claude-fable-5	5,858	35.6%	12.4%	52.0%	2.88
claude-opus-5	5,865	29.2%	9.7%	61.1%	3.02
deepseek-v4-pro	5,863	25.6%	18.4%	56.0%	1.39
gemini-3.1-pro	5,859	31.7%	20.0%	48.2%	1.58
gpt-5.6-sol	5,865	27.8%	22.8%	49.4%	1.22
grok-4.6 (spliced)	5,860	24.0%	22.2%	53.8%	1.08
qwen-3.8-max (spliced)	5,809	32.2%	19.5%	48.3%	1.65

L:S = each model's own long-share divided by its short-share; the field average of the per-model ratios is 1.83 this month.

'Sideways' was 52.7% of mature forecasts (21,596 of 40,979). Consensus skew: 2,225 symbol/FH/TF cells had 5 or more models on the same side and hit 45.3% (n=13,776) against the 46.3% field base.

Model	Symbol	Side	Conf	FH / TF	Slot (UTC)	Exit	Net PnL
gemini-3.1-pro	XRP	Long	85	4h / 1h	Aug 21, 16:01	SL	-\$3.10

Fail of the month. The month's single highest-confidence individual miss, exit reason sl. Listed as a single card, never as a model ranking.

Calibration bridge

claude-opus-5 is both this month's hit-rate leader (47.8%) and the best-calibrated line (Brier 0.2677, gap +13.2pp). Per-model ok-rate this window ran 99.06% to 100.00%. The full confidence-bucket analysis and the week-by-week gap drift live in Monthly Calibration #1.

Asset league: which coin the field reads best

A published week gives each coin 644 to 1,118 calls; the month gives 3,695 to 4,068 per coin. Across the month the field's hit rate ran from 48.1% on BNB down to 42.9% on ETH, a spread of +5.2pp against a 46.3% field base. The ordering below is the month pull; the third table shows how often that ordering held from week to week.

Symbol	N	Hit rate	95% CI	Rank
BNB	3,857	48.1%	46.5%-49.6%	1
SOL	3,928	47.3%	45.8%-48.9%	2
XRP	3,835	46.7%	45.1%-48.3%	3
BTC	4,068	46.5%	45.0%-48.0%	4
ETH	3,695	42.9%	41.3%-44.5%	5

Ranked by hit rate, ties broken by n descending; Wilson 95% CIs are descriptive and 6 of the 10 pairwise CI comparisons in this table overlap. Source: the month pull split by slot week, at the monthly cutoff.

Symbol	1h	4h	1d	Best FH
BNB	48.1% (2,364)	49.4% (1,244)	41.4% (249)	4h
SOL	45.3% (2,440)	49.0% (1,266)	59.9% (222)	1d
XRP	46.4% (2,427)	47.2% (1,166)	47.5% (242)	1d
BTC	46.5% (2,502)	44.8% (1,327)	55.6% (239)	1d
ETH	44.3% (2,377)	39.7% (1,122)	43.4% (196)	1h

Cells are hit rate with n in parentheses; * = N<10 (insufficient) and such cells never win the Best FH column, which is left n/a when no horizon clears the floor. Same pool and cutoff as the table above.

Symbol	W1 Aug 3-9	W2 Aug 10-16	W3 Aug 17-23	W4 Aug 24-30	Month
BNB	43.6% (3)	49.3% (1)	54.0% (3)	41.2% (4)	48.1% (1)
SOL	48.2% (1)	38.9% (4)	54.3% (2)	45.8% (1)	47.3% (2)
XRP	46.5% (2)	48.5% (2)	51.2% (5)	40.3% (5)	46.7% (3)
BTC	38.4% (4)	44.0% (3)	57.5% (1)	43.7% (2)	46.5% (4)
ETH	34.8% (5)	35.6% (5)	52.8% (4)	43.3% (3)	42.9% (5)

Cells are hit rate with the week's rank in parentheses. The four weekly columns are the PUBLISHED weekly numbers -- each week at its own frozen cutoff; the Month column is the single month pull. SOL ranked first in 2 of 4 weeks and 2 for the month; best horizon 1d (59.9% on n=3,928).

Rank stability across the four weeks

Every model's rank in each of the four published weeks, its rank for the pooled month, its mean rank and the span between its best and worst week, with the hit rate next to every rank. Mean rank runs 1.75 to 5.50, the widest span is 6 rank positions and the narrowest 2, and 3 distinct model lines held the weekly top spot across the four weeks.

Model	W1	W2	W3	W4	Month	Mean rank	Range
claude-opus-5	7 (40.9%)	7 (41.5%)	1 (59.1%)	6 (42.4%)	1 (47.8%)	5.25	6
gemini-3.1-pro	1 (43.7%)	1 (47.0%)	3 (54.1%)	2 (44.2%)	2 (47.6%)	1.75	2
claude-fable-5	6 (41.4%)	6 (41.7%)	2 (56.5%)	3 (43.0%)	3 (46.7%)	4.25	4
gpt-5.6-sol	3 (43.0%)	2 (44.2%)	4 (53.6%)	4 (42.7%)	4 (45.9%)	3.25	2
qwen-3.8-max (spliced)	2 (43.4%)	3 (44.1%)	5 (52.3%)	5 (42.6%)	5 (45.6%)	3.75	3
deepseek-v4-pro	5 (41.6%)	5 (42.5%)	6 (51.6%)	1 (44.6%)	6 (45.5%)	4.25	5
grok-4.6 (spliced)	4 (42.2%)	4 (44.0%)	7 (51.4%)	7 (39.9%)	7 (45.3%)	5.50	3

Rank = position in that window's leaderboard, which the pack sorts by hit rate descending; hit rate in parentheses. Weeks are the published issues at their own frozen cutoffs (W1 Aug 3-9, W2 Aug 10-16, W3 Aug 17-23, W4 Aug 24-30); the Month column is the single month pull, so the month rank is not the mean of the four weekly ranks. Consecutive-week Spearman rho: W1-W2 0.96, W2-W3 -0.32, W3-W4 0.07 (ties get average ranks). 3 distinct model lines held the weekly top spot. Sorted by month rank.

Calm vs volatile weeks

The pack labels each of the four weeks from the BTC daily series: a week is volatile when its mean absolute BTC 1d move is at or above the median of the four weeks, else calm, against a median mean |BTC 1d move| of 0.867% across the four weeks; the trend/flat day counts come from the same BTC series under the daily rule (a day is a trend day when its absolute BTC 1d move is 1.5% or more, else a flat day). 2 weeks come out calm and 2 weeks volatile, which splits the month into two regime cells.

Week	Mean BTC 1d move	Trend days	Flat days	Label
W1 Aug 3-9	0.529%	0	7	calm
W2 Aug 10-16	0.507%	1	6	calm
W3 Aug 17-23	3.610%	5	2	volatile
W4 Aug 24-30	1.206%	3	4	volatile

Mean |BTC 1d move| = the mean absolute daily BTC move of that week, from the entry-price series the pack reconstructs; the label is the pack's rule against the four-week median (0.867%). BTC only: this is not the market check's "Mean |1d move| (5 assets)" row, which averages five coins over the whole window.

Model	Calm n	Calm hit	Volatile n	Volatile hit	Diff pp
claude-opus-5	826	41.2%	1,233	51.5%	+10.3
claude-fable-5	1,114	41.6%	1,430	50.3%	+8.8
deepseek-v4-pro	831	42.1%	1,506	48.2%	+6.1
gpt-5.6-sol	1,248	43.6%	1,417	48.0%	+4.4
gemini-3.1-pro	1,184	45.3%	1,553	49.3%	+4.1
qwen-3.8-max (spliced)	1,282	43.7%	1,378	47.5%	+3.8
grok-4.6 (spliced)	1,227	43.1%	1,189	46.8%	+3.7
Field (all models)	7,712	43.1%	9,706	48.8%	+5.7

Definition: diff_pp = volatile hit_rate - calm hit_rate, in percentage points. Sorted by Diff pp descending, Field row last. * = N<10 (insufficient). Rows outside the four Mon-Sun weeks (1,965 calls on Aug 1-2 and Aug 31) are not in either regime cell, so the two columns do not sum to the month.

Market check: August ran hotter than the partial July

OBSERVATION — mean |1d move| 1.54% -> 1.98% (+29% rel), BTC realized vol 33.5% -> 36.9% (ann., hourly); field directional accuracy 45.0% -> 46.3% (+1.4pp), 5 of 7 models improved, sim win-rate up for 7 of 7, field sim PnL -\$3095.02 -> -\$1054.79 (Jul 11-31 (partial) vs Aug 1-31; descriptive, one pair of windows, not a claim). Robustness: raw price-sign accuracy 46.8% -> 46.0% — the same direction as the trade-based hit rule.

Measure	Jul 11-31 (partial)	Aug 1-31	Change
Mean 1d move (5 assets)	1.54%	1.98%	+29% rel
BTC realized vol (ann., hourly)	33.5%	36.9%	+3.5pp
Field directional accuracy	45.0%	46.3%	+1.4pp
Models improving hit-rate	--	5 of 7	--
Raw price-sign accuracy	46.8%	46.0%	-0.9pp
Field sim win-rate	31.9%	35.8%	+3.9pp
Field sim net PnL	-\$3095.02	-\$1054.79	--

Window A is the pack's prior window -- a PARTIAL July (forecasts from Jul 11), not a full month -- so this slice and every month-over-month column in this report compare a full August against a partial July. 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: grok-4.6 (incl. 4.5-era, Aug 1-24); qwen-3.8-max (incl. 3.7-era, Aug 1-5). 5,390 rows were remapped inside the month against 19,532 in the prior window.

Practical implications

- On this month's leaderboard no title was awarded, the top two CIs overlap, and 3 distinct model lines led the four weeks inside the same month.
- BNB ranked #1 for the month on $n=3,857$; the week-by-week league shows it ranked #1 in 1 of the four published weeks.
- The field differs by +5.7pp between the two regimes on $n=7,712$ calm vs $n=9,706$ volatile; the rule that draws the line is the pack's own four-week median, recomputed every month.

Limitations

- The prior window is a PARTIAL month: forecasts run only from Jul 11 and the audited daily candles only from Jul 15, so every "prior" column, the gray series of the page-1 chart and the market-check A column describe Jul 11-31 (partial), not a full July. It is window A of the same pull, slots 2026-07-01 to 2026-08-01, forecasts counted from Jul 11 (first slot actually seen 2026-07-15 18:01 UTC) and audited candles from Jul 15, matured at the prior cutoff Sat Aug 1, 2026, 16:00 UTC under the same engine, the same FH gate and the same hit rule as August -- 19,541 directional calls against 19,383 in the month. Month-over-month deltas are a step between two windows of different length.
- Two cutoffs live in this issue. The month is frozen at Mon Sep 1, 2026, 16:00 UTC; the four weekly blocks are frozen at their own published cutoffs (W1 Aug 3-9, W2 Aug 10-16, W3 Aug 17-23, W4 Aug 24-30). 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 (19,383 vs 17,418 directional, plus 1,965 on the 3 remainder days) -- that gap is arithmetic, not a data problem.
- Observations inside one window are not independent, and the 95% Wilson CIs shown throughout are descriptive, not inferential. Ranks are ordinal: a 0.3pp gap and a 6pp gap both read as one rank position, so the rank-stability table prints the hit rate next to every rank.
- The reversal detector remains v1 and BTC/ETH-only, on price series reconstructed from trade entry prices; 9 events in this window still cannot characterise detector performance. Cross-TF and cross-FH disagreement cells are small ($n=91, 23, 122, 32$); cells under $N=10$ are insufficient and never used to rank models.
- Model lines in this window: claude-fable-5, claude-opus-5, deepseek-v4-pro, gemini-3.1-pro, gpt-5.6-sol, grok-4.6 (spliced), qwen-3.8-max (spliced). Series density and lineage notes (monthly #1): (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 PARTIAL as measured -- the 1w series has slots on 13 of 31 days and the 1M series on 10 of 31 days; both sit outside this report's FH gate (1h/4h/1d) and appear only in the exclusions counter (1w 906, 1M 698). (2) 2 model lines are lineage splices inside this window: grok-4.6 (incl. 4.5-era, Aug 1-24); qwen-3.8-max (incl. 3.7-era, Aug 1-5). (3) Every per-model aggregate in this issue sees the merged head id; the raw id is kept in the pack, and 5,390 rows were remapped inside the month against 19,532 in 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 PARTIAL: audited daily candles start 2026-07-15, so it covers 17 days (Jul 15-31), and the TOP5 volume figure compares a 31-day sum with a 17-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 (Mon Sep 1, 2026, 16:00 UTC) -- the D2 definition pinned in the weekly series; this issue prints 39,073. 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 reproduces the published weekly #4 on 7 of 7 checks. Counter deltas against pre-wave-3 issues remain definitional.

Counters & lineage

Counter	Value	Counter	Value
Forecasts total	42,654	Mature (scored pool)	40,979
OK in gate	40,979	-- of them directional	19,383
Out of gate (1w / 1M)	906 / 698	-- of them sideways	21,596
Invalid	71	Pending (next issue)	0
Late closes	0	Uptime, all grids	1,173 / 1,178 slots
Source file	monthly_metrics_2026-08.json	Report cutoff	Mon Sep 1, 2026, 16:00 UTC
Generated at (pipeline)	2026-09-03 12:43 UTC	Methodology	v1.1 (2026-08-10) · hash e66c7e8c864a2233

THIS REPORT: 19,383 scored observations. MARKETMANIA RESEARCH TO DATE (as of cutoff Sep 1, 16:00 UTC): 39,073 resolved since Jul 11 · 7 models tracked (current line-up; earlier versions folded into their successors' lineage) · 5 assets · 5 horizons · hourly · 21 published reports.

Issue #1. Monthly Model Watch is a living monthly comparison; each issue appends one more calendar month of leaderboard, asset-league, rank-stability and regime data. This first issue sets the baseline: claude-opus-5 at the top (47.8%), 6 of 9 qualifying turns found a caller, 0 of 5 tickers above 51%, 3 distinct model lines leading the four weeks and a calm-vs-volatile field difference of +5.7pp. 2 model lines are lineage splices this month (grok-4.6 (incl. 4.5-era, Aug 1-24); qwen-3.8-max (incl. 3.7-era, Aug 1-5)). 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

- Monthly #2 (September): the same three series over Sep 1-30, cut on Oct 1 plus the maturity lag, so the first month-over-month comparison runs against a FULL prior month instead of a partial one.
- Monthly Config Watch #1: the same August window through the config search engine (w200), published as the fourth monthly report of this issue.
- Quarterly Stability Index: the next identical-prompt run lands around Oct 8 and is the first cross-quarter point on model stability.

Related research

Report	Direct PDF link
Monthly Consensus Watch #1	https://marketmania.ai/research/reports/consensus-watch-monthly-2026-08.pdf
Monthly Calibration #1	https://marketmania.ai/research/reports/calibration-monthly-2026-08.pdf
Weekly Model Watch #4	https://marketmania.ai/research/reports/model-watch-2026-08-24.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_modelwatch_monthly_2026m08,
  title = {Monthly Model Watch #1: monthly leaderboard, asset league, rank stability and regime split, Aug 1-31 2026},
  author = {{MarketMania Research}},
  year  = {2026}, month = {September}, day = {4},
  url   = {https://marketmania.ai/research/reports/model-watch-monthly-2026-08.pdf},
  note  = {Methodology v1.1 (2026-08-10), hash e66c7e8c864a2233; window Aug 1-31, 2026 UTC; source monthly_metrics_2026-08.json}
}
```