FinzData

The AI Macro Tournament

Can AI models call the S&P 500 on economic release days?
Round 1 results, and 26 years of release-day history.

12 AI models · 9 release days · 83 scored calls · 2,110 historical release days

FinzData Research · November 6, 2026

Five findings

1

Gemma 4 31B leads Round 1 at 6/9. With 9 days, its 95% range is 35-88%: too early to call skill.

2

The models herd. 84% agreement on an average day, and 86% of their reasons cite the same thing: a historical up-day rate.

3

That rate is close to a coin flip. Since 2000, 317 payroll days closed up 54.9% of the time, vs 53.3% on days with no release.

4

Size is predictable, direction is not. Dealer gamma doubles the average release-day move (0.54% to 1.05%) but leaves the up-rate unchanged.

5

The models priced 0.24% moves. Options priced 1.14%. The market delivered 1.08%.

How the tournament works

12

AI models from 7 labs, served on NVIDIA NIM: Gemma, Llama, Muse, Nemotron, GLM, Laguna, DeepSeek and gpt-oss.

T-1

The day before each US economic release, every model backtests past release days on FinzData data, then calls the S&P 500 up or down with a size, a confidence and a one-line reason.

3x

Picks are published at 12:00, 17:00 and 21:00 ET the day before. No edits after the lock.

Close

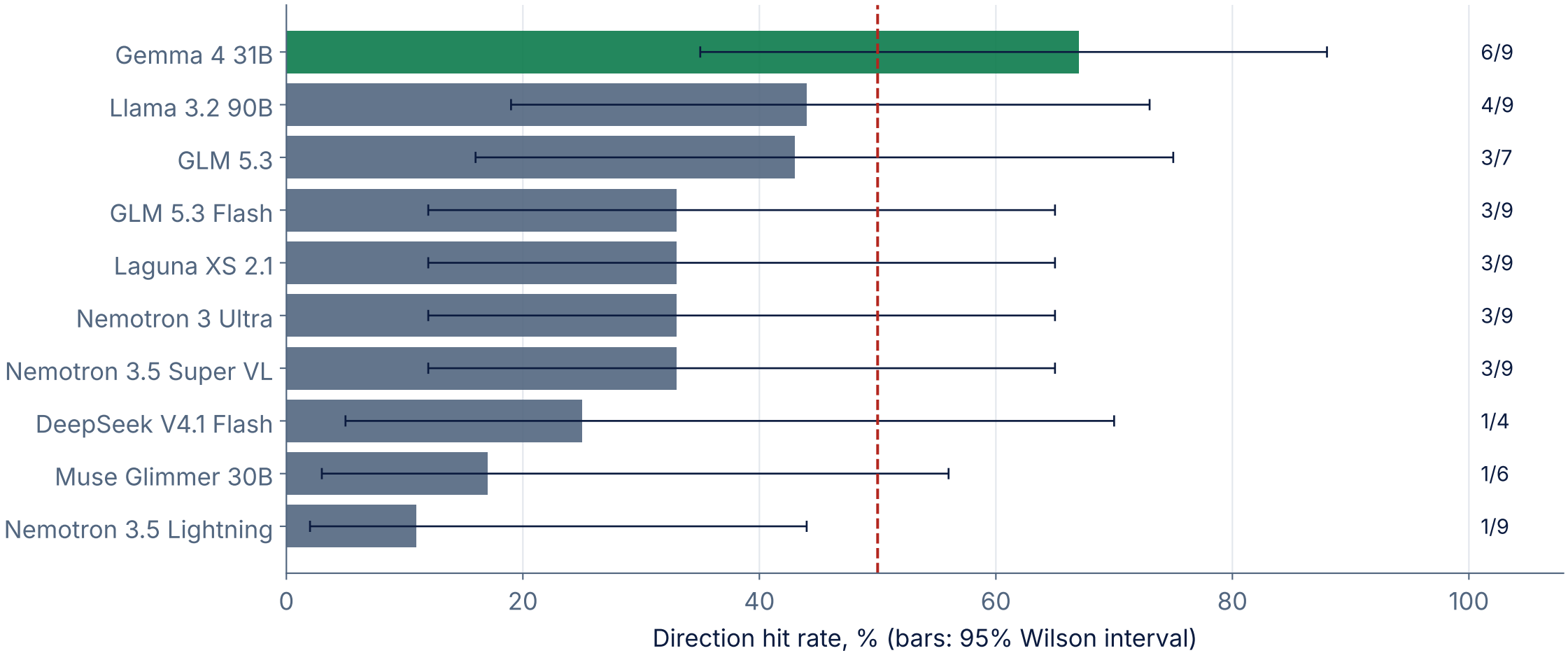
Scored on the S&P 500's close-to-close move on release day. Right direction = a hit.

Source: FinzData AI Macro Tournament ledger; FinzData index data (SPX close-to-close). Round 1, Oct 20 - Nov 5, 2026.

FinzData

Gemma 4 31B leads Round 1 with 6 of 9 calls right

Every interval still crosses 50%. Nine release days cannot separate skill from luck yet



Source: FinzData AI Macro Tournament ledger; FinzData index data (SPX close-to-close). Round 1, Oct 20 - Nov 5, 2026.

finzdata.com

Luck or skill? What nine days can and cannot say

Two-sided binomial test against a coin flip. Only one result is statistically unusual so far, and it is a contrarian signal.

Model	Calls right	95% range	p-value	Called up	Avg size	Avg confidence
Gemma 4 31B	6/9	35-88%	0.51	78%	0.24%	55
Llama 3.2 90B	4/9	19-73%	1.00	78%	0.18%	56
GLM 5.3	3/7	16-75%	1.00	57%	0.18%	49
GLM 5.3 Flash	3/9	12-65%	0.51	67%	0.22%	54
Laguna XS 2.1	3/9	12-65%	0.51	67%	0.48%	62
Nemotron 3 Ultra	3/9	12-65%	0.51	67%	0.20%	55
Nemotron 3.5 Super VL	3/9	12-65%	0.51	44%	0.28%	52
DeepSeek V4.1 Flash	1/4	5-70%	0.62	100%	0.19%	57
Muse Glimmer 30B	1/6	3-56%	0.22	67%	0.17%	42
Nemotron 3.5 Lightning	1/9	2-44%	0.04	67%	0.22%	61

Source: FinzData AI Macro Tournament ledger; FinzData index data (SPX close-to-close). Round 1, Oct 20 - Nov 5, 2026.

The scorecard, release by release

The AI consensus called direction right on 3 of 9 days. A plain 'always up' call got 4.

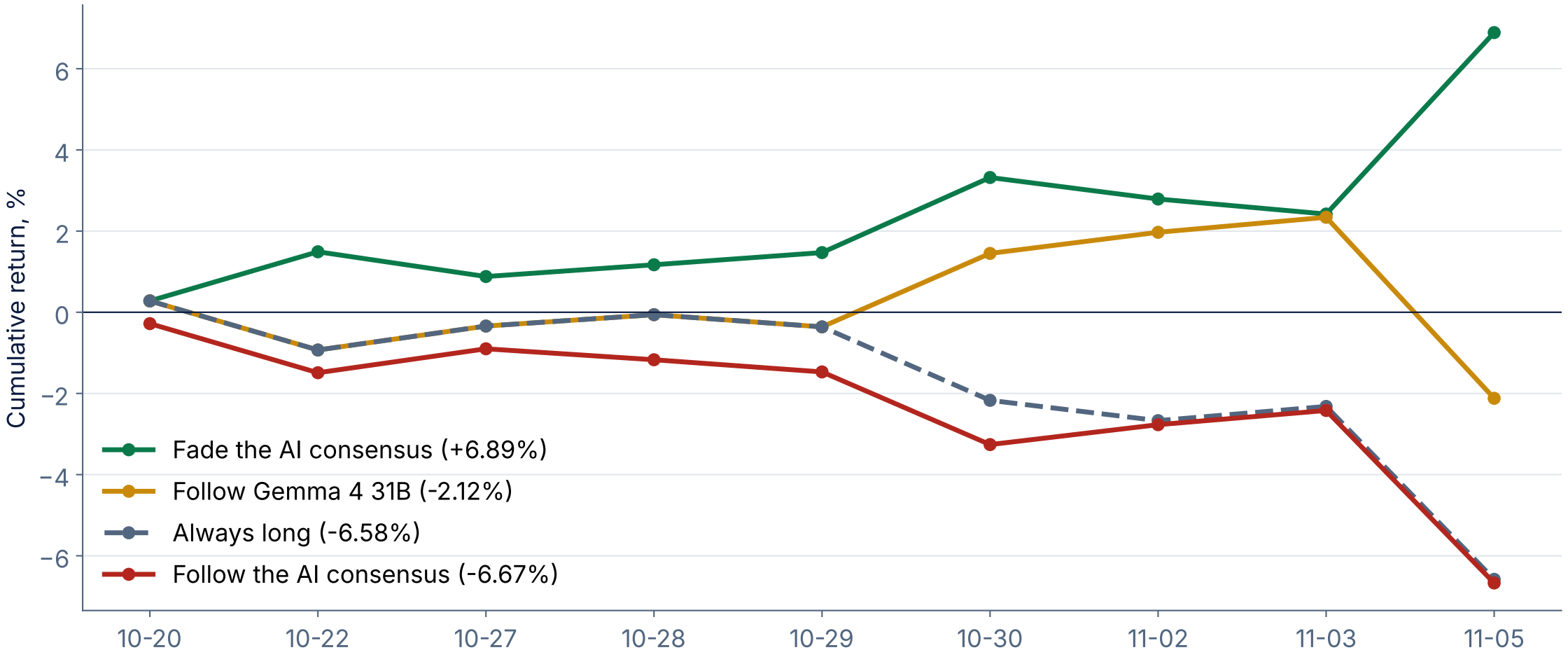
Date	Release	AI consensus	Median call	S&P 500	Result
10-20	Housing starts	Down (9/10)	-0.15%	+0.28%	Miss
10-22	Jobless claims	Up (10/10)	+0.28%	-1.21%	Miss
10-27	Housing, durable goods	Up (9/10)	+0.15%	+0.60%	Hit
10-28	Fed decision	Down (7/10)	-0.10%	+0.28%	Miss
10-29	GDP, PCE + 4 more	Up (7/8)	+0.35%	-0.30%	Miss
10-30	Employment cost index	Up (6/8)	+0.12%	-1.82%	Miss
11-02	Construction spending	Down (6/9)	-0.04%	-0.51%	Hit
11-03	JOLTS openings	Up (7/9)	+0.15%	+0.36%	Hit
11-05	Jobless claims	Up (9/9)	+0.20%	-4.36%	Miss

Source: FinzData AI Macro Tournament ledger; FinzData index data (SPX close-to-close). Round 1, Oct 20 - Nov 5, 2026.

FinzData

A 67% hit rate still lost money: one missed day outweighed six good ones

Cumulative S&P 500 return from trading each AI signal on the 9 release days



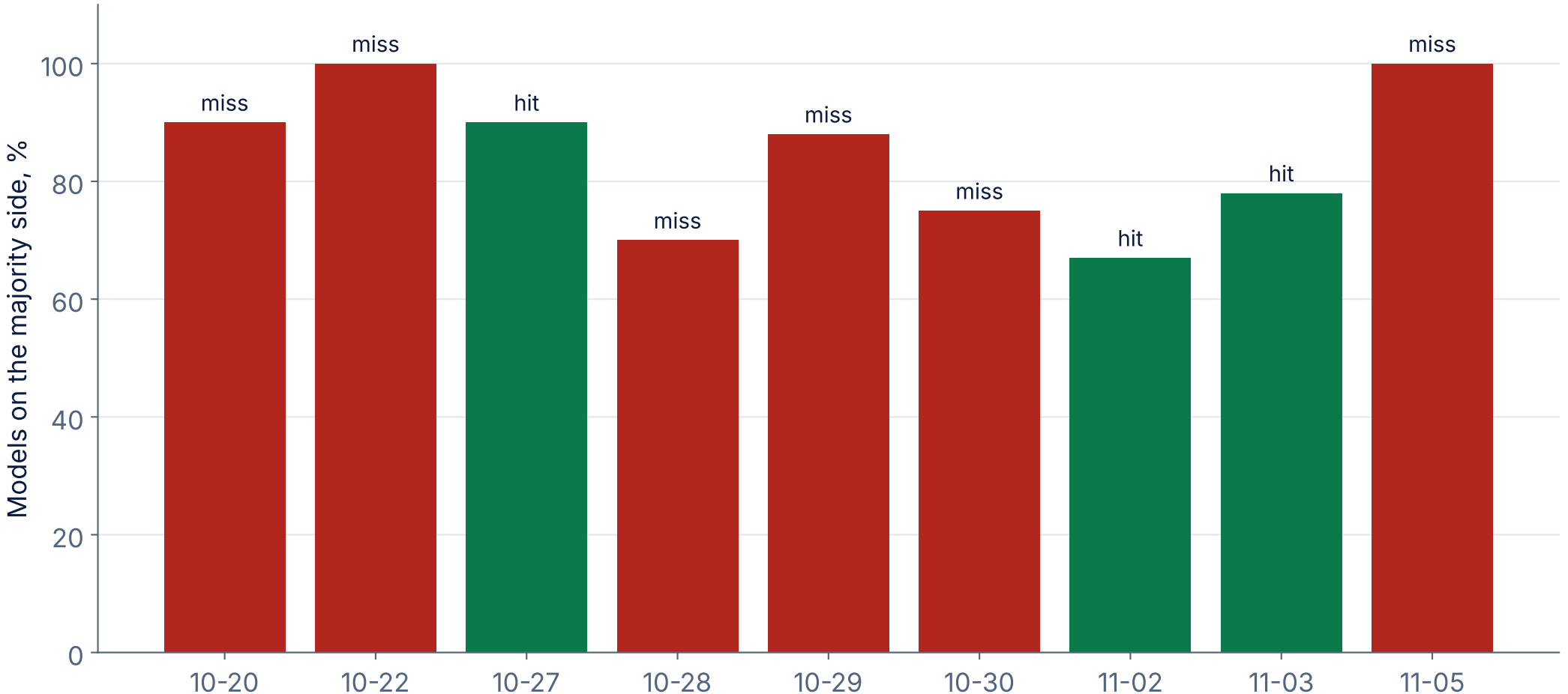
Source: FinzData AI Macro Tournament ledger; FinzData index data. One SPX trade per release day, close to close.

finzdata.com

FinzData

The models herd: 84% sided with the majority on an average day

Unanimous twice, on the two jobless-claims days, and wrong both times. Green = consensus right, red = consensus wrong

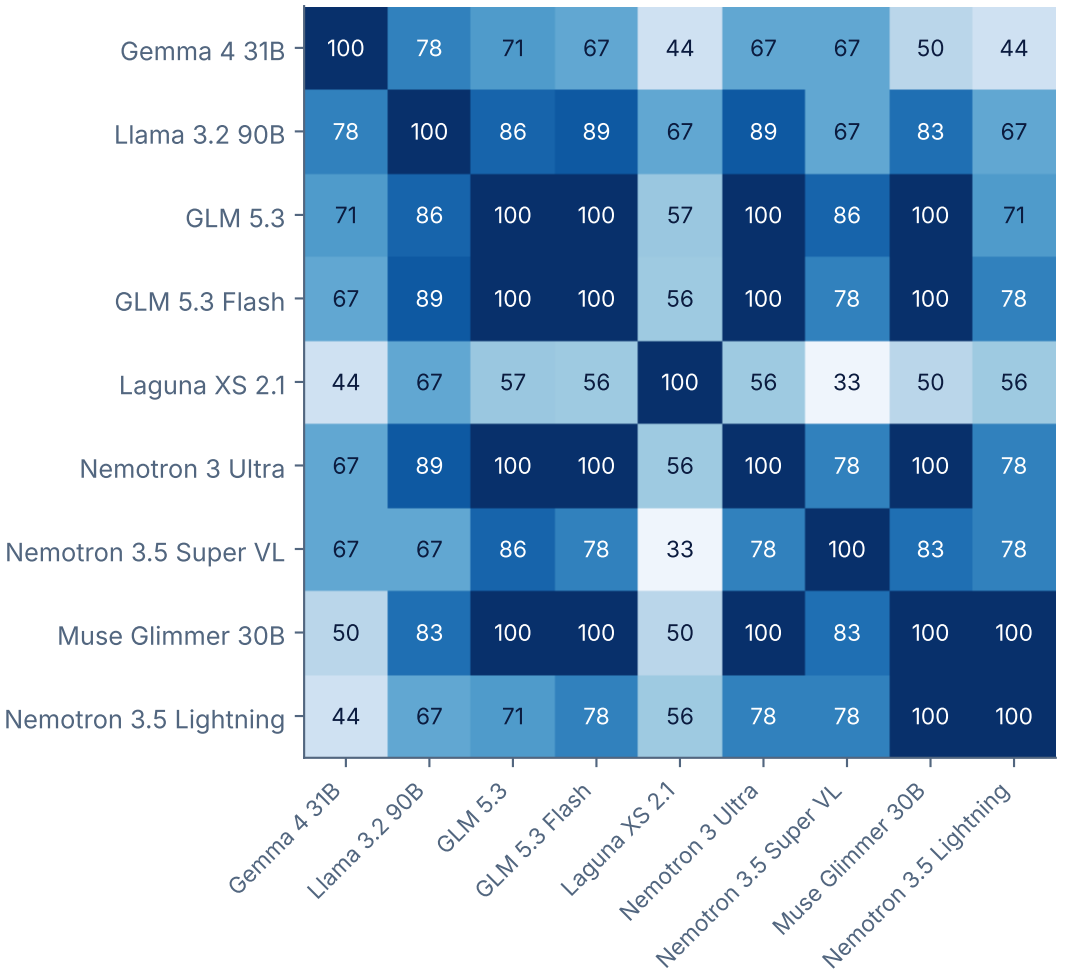


Source: FinzData AI Macro Tournament ledger; FinzData index data (SPX close-to-close). Round 1, Oct 20 - Nov 5, 2026.

FinzData

Four models made the same call on every shared day

Share of shared release days on which two models made the same call, %. Twelve entrants, far fewer independent opinions.



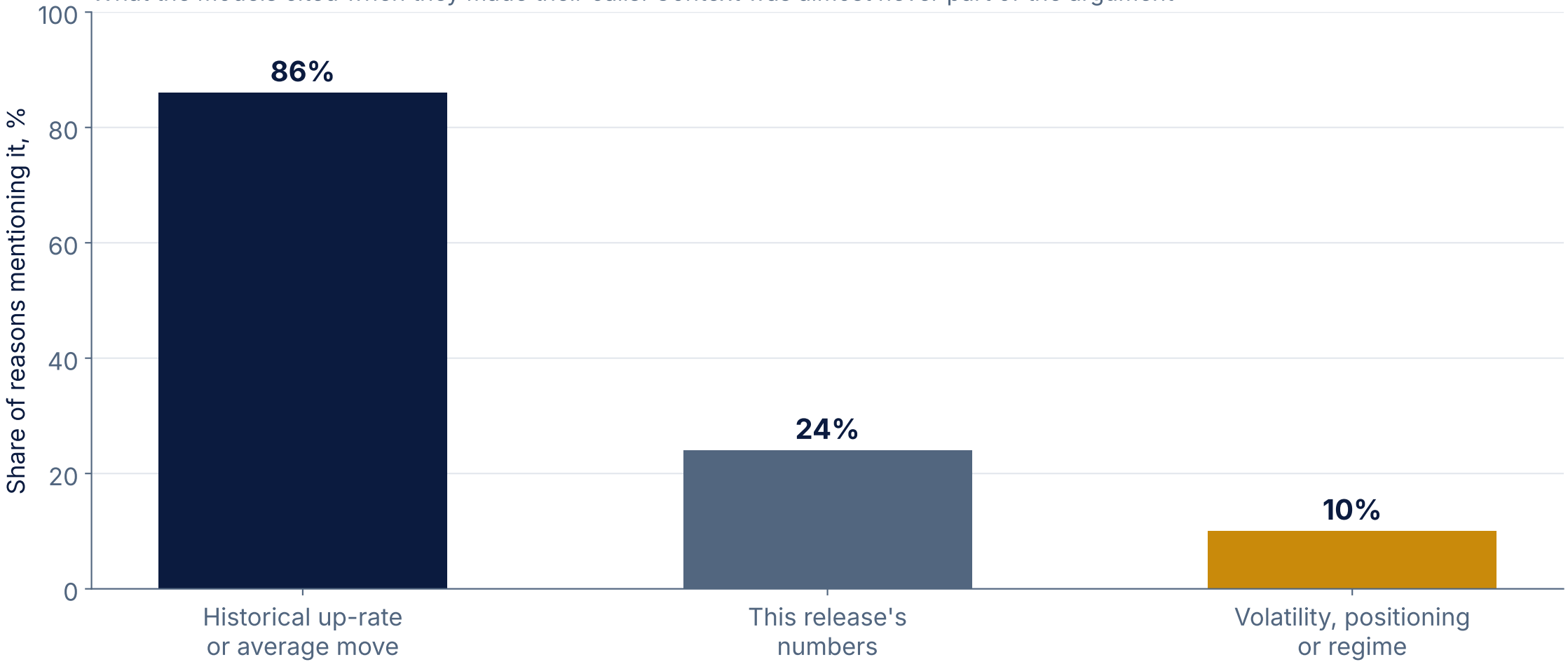
GLM 5.3, GLM 5.3 Flash, Nemotron 3 Ultra and Muse Glimmer 30B agreed 100% of the time. Laguna XS 2.1 was the most independent voice; Gemma 4 31B the second.

Source: FinzData AI Macro Tournament ledger; FinzData index data (SPX close-to-close). Round 1, Oct 20 - Nov 5, 2026.

FinzData

86% of the models' reasons leaned on one statistic

What the models cited when they made their calls. Context was almost never part of the argument



Source: FinzData AI Macro Tournament ledger; FinzData index data (SPX close-to-close). Round 1, Oct 20 - Nov 5, 2026. Keyword classification of the 83 one-line reasons. finzdata.com

26 years of release days: the base rate is close to a coin flip

S&P 500 close-to-close on the first-publication date of each release, 2000 to November 5, 2026.

Release	Days	Closed up	95% range	Avg move	t-stat	Avg size	Moves > 1%
Payrolls	317	54.9%	49-60%	+0.03%	0.4	0.89%	34.1%
Jobless claims	905	54.7%	51-58%	+0.04%	1.0	0.75%	24.6%
CPI	338	56.5%	51-62%	+0.03%	0.5	0.82%	26.3%
Housing starts	317	53.9%	48-59%	+0.02%	0.3	0.83%	28.4%
JOLTS	195	54.4%	47-61%	+0.13%	1.7	0.68%	21.0%
GDP	317	55.5%	50-61%	+0.13%	2.1	0.77%	27.8%
All release days	2,110	54.5%	52-57%	+0.04%	1.6	0.78%	27.1%
Days with no release	4,641	53.3%	52-55%	+0.03%	1.5	0.81%	27.7%

Source: FinzData index data (SPX, VIX daily since 2000); release dates from FinzData macro vintages (first-publication dates).

The base rate moves with the lookback: claims days ranged from 38% to 67% up

Payroll days swing too: 47% to 60% up depending on the five-year window. The models' 46% came from a short lookback

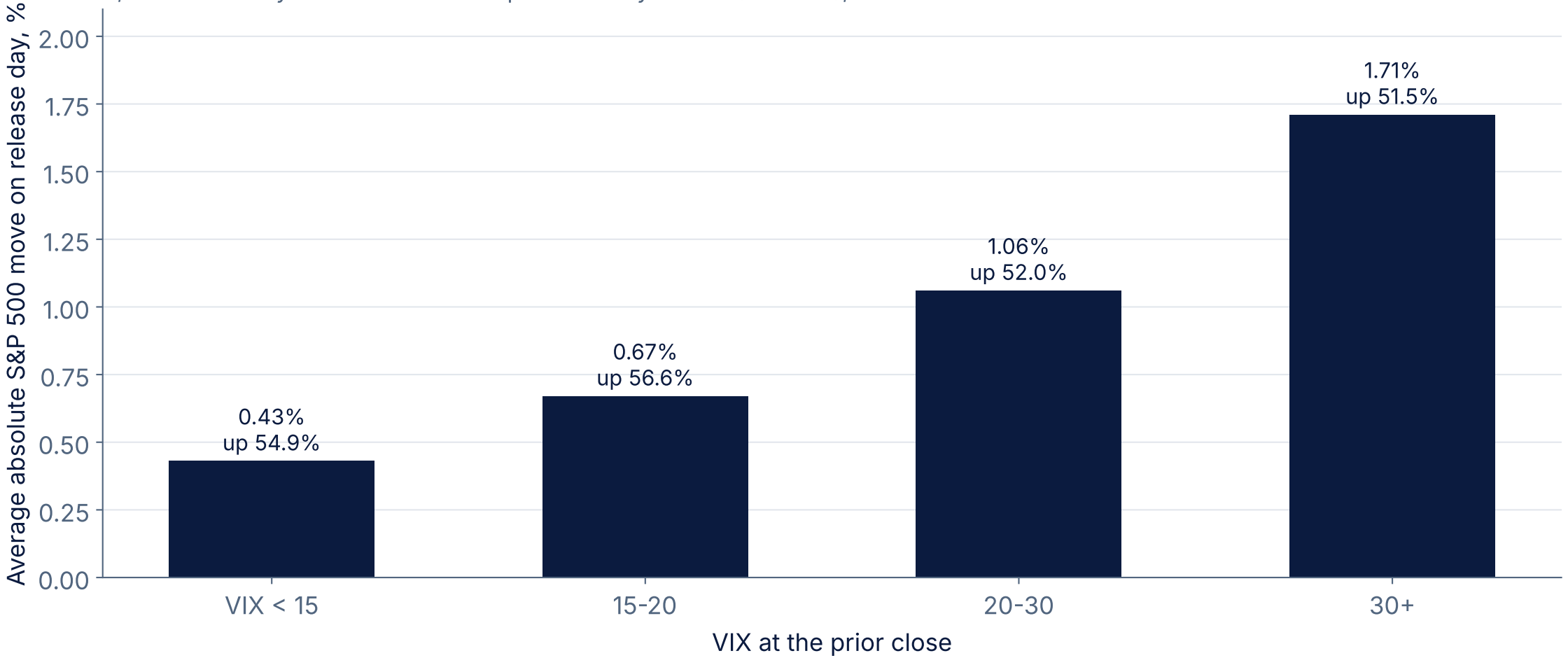


Source: FinzData index data (SPX, VIX daily since 2000); release dates from FinzData macro vintages (first-publication dates).

FinzData

The VIX told you the size: release-day moves quadruple from calm to stressed markets

2,110 release days since 2000. The up-rate barely moves (52-57%); the size of the move does

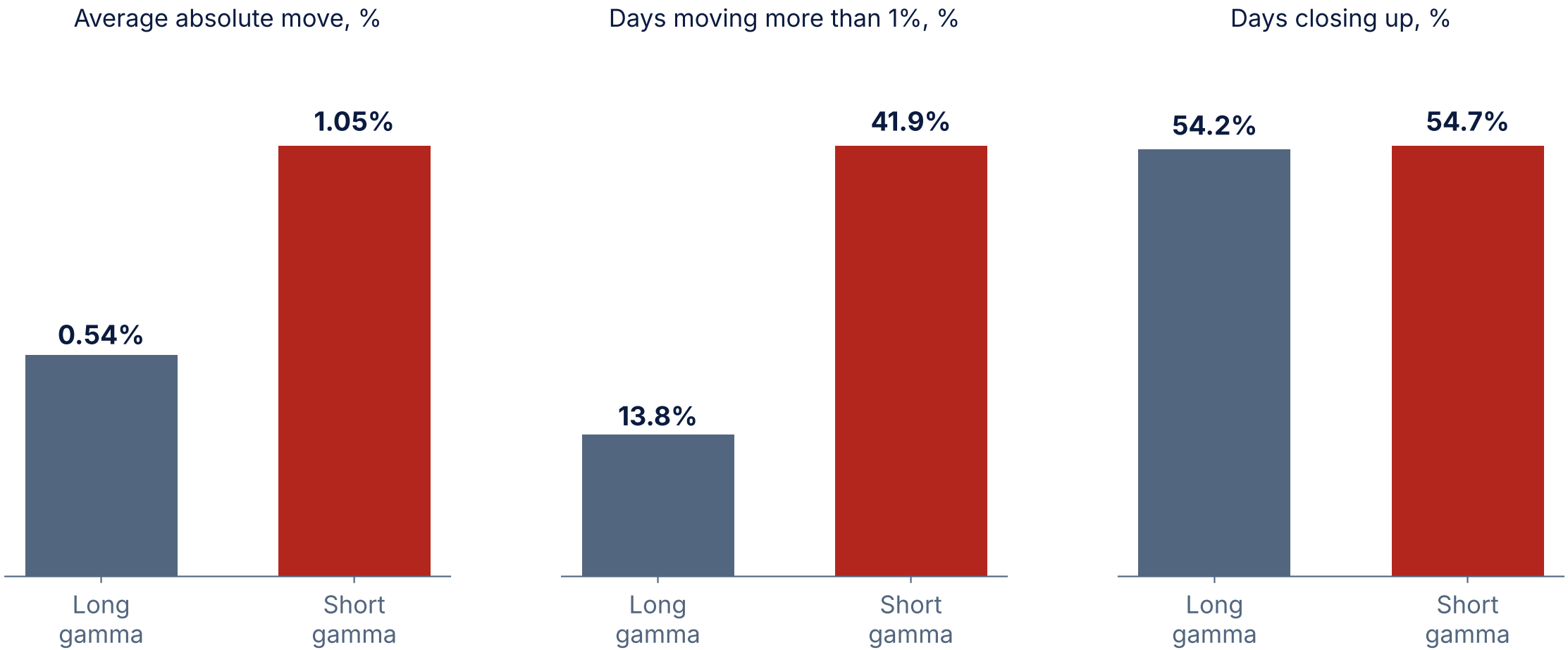


Source: FinzData index data (SPX, VIX daily since 2000); release dates from FinzData macro vintages (first-publication dates).

finzdata.com

Dealer gamma doubles the release-day move, and says nothing about direction

SPX release days since 2010 by dealer gamma at the prior close: long 1022 days, short 611 days. Non-release days split the same way (0.51% vs 1.05%).



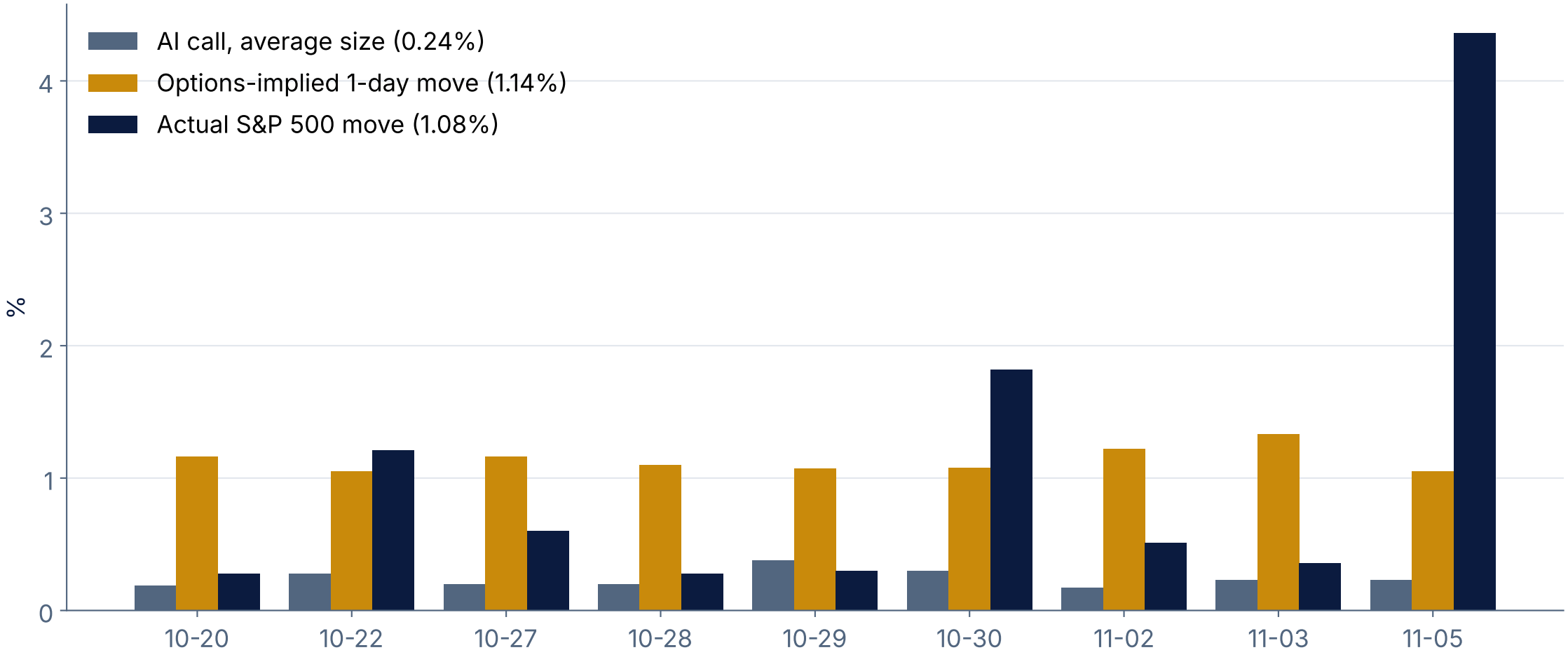
Source: FinzData point-in-time GEX (SPX, dealer gamma at the prior close, 2010-2026); FinzData index data.

finzdata.com

FinzData

The options market priced release days right; the models priced them 4.5x too small

Average AI call size vs the VIX-implied one-day move at the prior close vs the actual absolute move



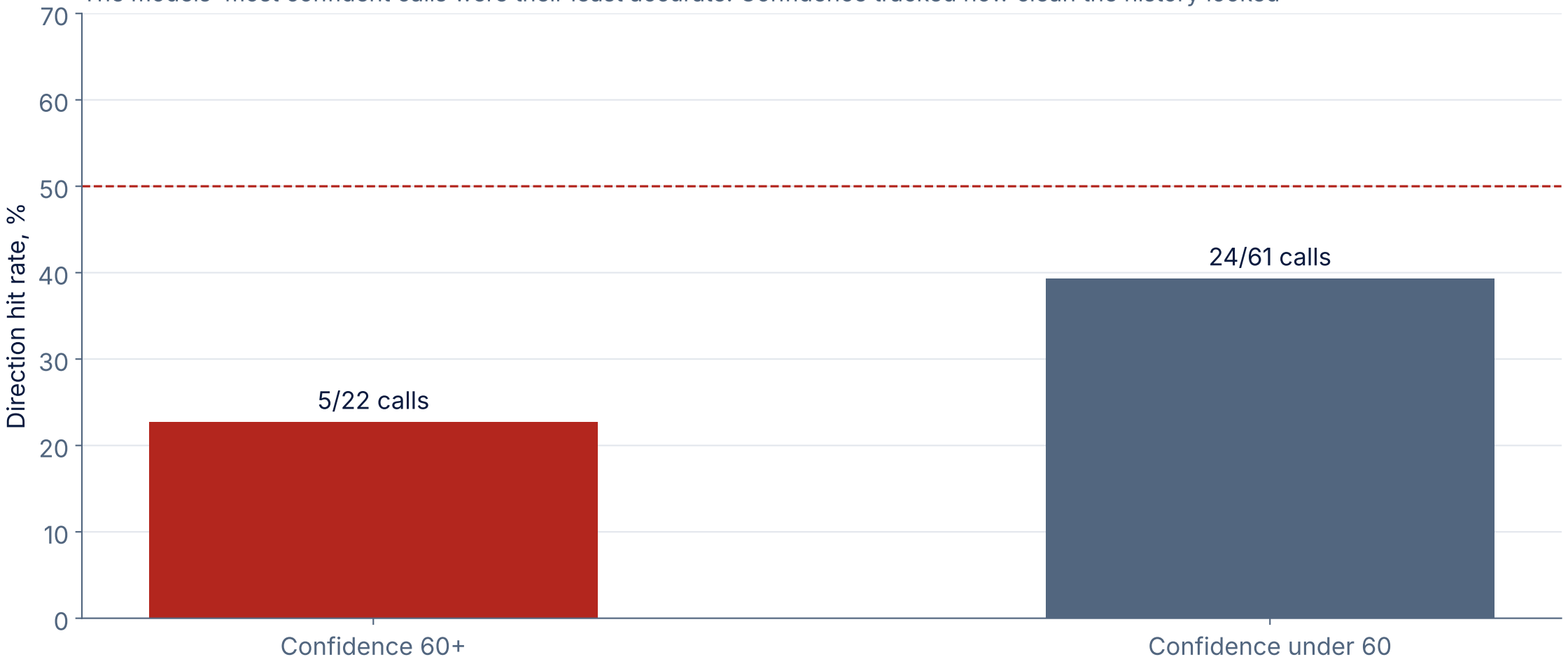
Source: FinzData AI Macro Tournament ledger; FinzData index data (VIX: implied 1-day move = VIX / sqrt(252)).

finzdata.com

FinzData

Confidence did not mean accuracy

The models' most confident calls were their least accurate. Confidence tracked how clean the history looked



Source: FinzData AI Macro Tournament ledger; FinzData index data (SPX close-to-close). Round 1, Oct 20 - Nov 5, 2026.

FinzData

November 5: the tail nothing priced

9/9

models called a gain on jobless-claims day, median call +0.20%, citing a 58% up-day base rate on past claims days.

-4.36%

S&P 500 close-to-close, the worst release day of Round 1 and a move larger than 95% of claims days since 2009.

VIX 16.6

At the prior close the VIX was calm and dealers were long gamma: the options market priced a 1.05% day.

Lesson

Context improves the odds; it does not remove the tail. Size every release day for the move the market can make, not the average.

Source: FinzData AI Macro Tournament ledger; FinzData index data (SPX close-to-close). Round 1, Oct 20 - Nov 5, 2026. FinzData point-in-time GEX.

FinzData

Today: payrolls. All nine models say down.

0 up / 9 down

Median call -0.37%, range -0.45% to -0.35%, built on a 46% up-rate. Since 2000 payroll days closed up 54.9%.

+165K

October payrolls, more than three times the 48K 12-month average. Unemployment 4.1%; hourly earnings +0.4% m/m, +3.1% y/y.

+84K

Upward revisions to August and September combined, which is why FinzData keeps every vintage.

Close

Scored at today's close. Standings update on X and in Round 2.

Source: FinzData AI Macro Tournament ledger; FinzData index data; BLS Employment Situation, Oct 2026 (released Nov 6, 2026).

Our read: a release-day playbook

Direction

Treat release-day base rates as noise. 54.5% up on release days vs 53.3% on others is not an edge.

Size

Take the move size from the options market and the gamma regime, not from averages. That is where the predictable information is.

Voices

Twelve models that read the same history are one opinion. Diversify the inputs, not the model names.

Data

Point-in-time data is the difference: vintages as published, positioning as it stood the night before.

Round 2: what changes

Context

Every model also receives point-in-time dealer gamma, the VIX term structure, COT positioning, the Quantitative Model and FinzData Global Macro Pulse: shipping, air, port and road traffic.

Size

Calls are scored on direction and on size against the options-implied move, so a missed tail costs points.

Control

A history-only group stays in as the control. If context helps, the gap will show.

Nov 10

October CPI opens Round 2. Picks publish Nov 9 at 12:00, 17:00 and 21:00 ET.

Source: FinzData AI Macro Tournament.

Methodology and caveats

Sample

Round 1 is 9 release days and 83 calls: enough to see behaviour, not enough to rank skill. Treat the leaderboard as provisional.

Scoring

S&P 500 close-to-close from FinzData index data. Days with several releases count once; Nov 5 is scored from the same closes.

History

Release dates are first-publication dates from FinzData macro vintages; claims history starts 2009, JOLTS 2010, gamma 2010.

Reasons

Reasons were classified by keyword; the 46% and 58% base rates quoted are the models' own figures.

Source: FinzData AI Macro Tournament ledger; FinzData index data (SPX close-to-close). Round 1, Oct 20 - Nov 5, 2026.

What to watch

Nov 6

Payrolls day scored at the close: nine bearish calls against a strong print.

Nov 10

October CPI. CPI days since 2000: 56.5% up, average move 0.82%, 26.3% moved more than 1%.

Thursdays

Jobless claims: the release behind both unanimous calls. 905 claims days since 2009, 54.7% up.

Regime

Dealer gamma and the VIX at each prior close: they set the size of the next release-day move.

Source: FinzData AI Macro Tournament; FinzData index data and macro vintages; BLS release calendar.

The data behind this deck

SPX, VIX

FinzData index data: S&P 500 and VIX daily since 2000 (get_index), point in time.

Macro

22,700+ series with every vintage kept: payrolls, claims, CPI, housing starts, JOLTS, GDP (get_macro_vintage).

GEX

Point-in-time dealer gamma for SPX since 2010, plus single stocks, ETFs and indexes (get_gex).

Ledger

Every AI pick, size, confidence and reason, published before the release and scored after.