Rillor

RESEARCH REPORT

What research shows about prediction market accuracy

Studies from 2004 to 2026 find prediction market prices broadly track outcomes, with a longshot bias, weaker calibration in political contracts and disputes over how accuracy is measured.



15 November 2026 Forecasting and market intelligence rillor.com/insights/prediction-market-accuracy-research

Rillor

Contents

01 Summary

02 Background

03 Method

04 Findings

05 Implications

06 Limits of this analysis

07 Sources

Summary

Research from 2004 to 2026 finds that prediction market prices broadly track outcomes, that cheap contracts win less often than their prices imply, and that each headline accuracy figure depends on what its study measured, on which contracts and how long before the outcome [5, 14, 18, 24, 28].

- Early studies compared markets with polls. Compared with 964 polls over the five U.S. presidential elections from 1988 to 2004, the Iowa Electronic Markets' vote-share forecast was closer to the outcome in 74% of the comparisons, and significantly better in every election when forecasting more than 100 days ahead [5].
- A price is an imperfect reading of what traders believe. Manski showed in 2004 that, under his assumptions, a price reveals nothing about how widely beliefs vary and only bounds their average, and Wolfers and Zitzewitz replied in 2006 that prices are usually close to traders' mean beliefs [9, 10].
- Cheap contracts win less often than their prices imply, a pattern called the favorite-longshot bias. In a 2026 analysis of 72.1 million Kalshi trades, contracts trading at 5 cents won 4.18% of the time [15], and in 2024 political markets, contracts priced at $0.10 to $0.20 averaged $0.149 and resolved Yes 7.1% of the time [28].
- In large samples, prices are calibrated on aggregate, meaning a contract priced at p resolves Yes about p percent of the time. A PNAS study of 19 October 2026 found this across about one million contracts and questions on four forecasting platforms, with systematic over- or underestimation in some highly consequential areas [18].
- Calibration varies with subject and horizon. A 2026 preprint on 429,000 contracts on Kalshi and Polymarket found political prices compressed toward 50% on both platforms, and prices overall more compressed the further a market was from resolution [17].
- The share of markets called right runs high partly because most contracts are lopsided. In a June 2026 study of 2024 political markets, 86% beat a coin flip, and 66% were priced below $0.20 or above $0.80, of which 96% resolved as priced, against 55% of markets priced between $0.40 and $0.60 [28].
- Accuracy rises as resolution nears. A platform-run Kalshi paper, as reported in August 2026, found the Brier score, the average squared error of the price against the outcome where lower is better, fell from roughly 0.08 to 0.09 three months out to about 0.02 at close [23, 24].
- In three studies of LLM forecasters from 2025 and 2026, scored by Brier score, the market price beat every agent on the 104 matches of the 2026 World Cup [33] and every system on one report's 1,610 liquid-market questions, though that report's system beat the market on 76 ForecastBench market questions, 0.0753 against 0.0965 [32]. In the third study, Prophet Arena, the market trailed the top model by 0.003 [31].

Background

Event contracts and their prices

The U.S. Commodity Futures Trading Commission (CFTC) describes event contracts as "often based on yes-no scenarios, allowing for only two possible outcomes", with "a fixed payout (usually $1) and an expiration" [1]. Its example is a contract at 70 cents for yes and 30 cents for no, a price that "reflects the market's expectation of the outcome" [1]. The agency writes that "A contract's price reflects traders' perceived probability of the event outcome", notes that event contracts "are typically structured as swaps", and says prediction markets in the United States "have been around since 1988 and regulated by the CFTC since 2004" [1].

Read as a probability, a 70-cent price, the CFTC's example [1], implies that about 70% of contracts priced like it resolve Yes. One 2026 study of Polymarket states that reading: under risk neutrality and ignoring the time value of money, a price is the market's implied probability, and if markets are efficient and carry no risk premium, contracts priced at p resolve Yes about p percent of the time [20]. Most of the studies below test that reading.

The platforms studied

The CFTC's timeline names the Iowa Electronic Markets (IEM), created in 1988 at the University of Iowa as an "experimental and academic program", as "the first modern prediction market" [1]. The IEM calls itself "a futures market run for research and teaching purposes", and its 2026 congressional markets were open when viewed [2]. A 2026 study lists its stakes as $5 to $500 per trader, and notes that PredictIt allowed 5,000 traders per market in 2024 and charged a 10% fee on profits [28].

Most recent research uses Kalshi and Polymarket data. Kalshi studies cover markets from the platform's 2021 launch [14, 23], and two 2026 Polymarket studies use trade data from November 2022 to March 2026 [16, 20]. A 2013 study used Intrade transaction prices [13], and a 2026 cross-platform study added Manifold and Metaculus [19].

How accuracy is scored

The studies use five measures, and each answers a different question.

- **Calibration** asks whether a contract priced at p "resolves in its favor approximately p percent of the time" [20]. It describes many contracts taken together and, in the words of one 2026 study, "does not mean that forecasting platforms provide certainty about individual events" [18].
- **The Brier score** is "the average squared error of predictions", in Polymarket's words [25]. Lower is better, and a constant 50% forecast scores 0.25 (derived).
- **Log skill** compares a market's log score with "an uninformed 50-50 baseline", so log skill above 0 beats a coin flip [26].

- **Naive accuracy** "treats the side priced above 50 cents as the predicted winner" and "scores a 51-cent contract that resolves Yes the same as a 95-cent contract that resolves Yes" [24].
- **The recalibration slope** fits outcomes to prices. A slope above 1 means "favorites are underpriced, longshots are overpriced and prices are compressed toward 50%" [17].

Method

What was examined. Published evidence on how well prediction market prices forecast outcomes: comparisons with polls, what a price can identify, the favorite-longshot bias, calibration in large samples, platform-run figures, the dispute over a 2024 election study, markets outside politics and market prices as a baseline for LLM forecasters.

Sources. Peer-reviewed articles from 1988 to 2026 [3, 5, 6, 11, 13, 18, 30], working papers and preprints [4, 7, 8, 9, 10, 12, 14, 16, 17, 19, 20, 26, 28, 29, 31, 32, 33], an independent researcher's web paper [15], regulator, university and platform pages [1, 2, 21, 22, 25] and three news reports [23, 24, 27].

How. Texts were read in full where they could be opened, and each figure was taken from the passage or table cited. Five sources were read from their published abstracts [3, 6, 8, 11, 29]. The Kalshi Research working paper could not be opened, so its figures come from two news reports and are labelled platform-run [23, 24]. Only the published abstract of the PNAS article was read, from its DOI record, so the counts and examples for that study come from the same team's March 2026 preprint [18, 19]. Platform pages were viewed on 8 November 2026, Polymarket's again on 14 November 2026, and their figures are as viewed. Each figure stays with its own measure, sample and date, none is averaged across studies, and no platform is ranked, since the authors of the 2024 election study write that summary accuracy is "not a statement about the relative quality of the various exchanges" [28]. Return figures from the longshot studies are reported only as the measured size of the bias. Figures marked "derived" are arithmetic on a source's own numbers.

Findings

1. Early evidence: markets against polls

Wolfers and Zitzewitz reported in 2004 that across the four U.S. presidential elections from 1988 to 2000, the IEM's vote-share contracts predicted the two major parties' vote shares "with an average absolute error of around 1½ percentage points" in the final week, while the final Gallup poll "erred by 2.1 percentage points" [4]. They added that "the accuracy of the market prediction improves as information is revealed and absorbed as the election draws closer" [4]. The published version concludes that "market-generated forecasts are typically fairly accurate, and that they outperform most moderately sophisticated benchmarks" [3].

Berg, Nelson and Rietz moved the test away from election eve in 2008. They compared the IEM's vote-share market with 964 polls over the five presidential elections from 1988 to 2004 [5]. The market was closer to the eventual outcome in 74% of those comparisons and "significantly outperforms the polls in every election when forecasting more than 100 days in advance" [5].

The record runs further back. Rhode and Strumpf studied presidential betting markets "that operated between 1868 and 1940" and found that "the market did a remarkable job forecasting elections in an era before scientific polling" [6].

Sethi and colleagues set a published election model from The Economist against PredictIt prices across the 2020 battleground states, with seven months of daily data and mean Brier scores [7]. The market did better several months out and the model as the election approached, and "A simple average of the two forecasts performs better than either one of them overall" [7]. For 2024, a July 2025 preprint concluded that Polymarket "was superior to polling in predicting the outcome of the 2024 presidential election, particularly in swing states", while calling for further work on its "intriguing results" [8].

2. What a price can and cannot reveal about beliefs

Traders can hold different beliefs, so a price is a summary of them. Manski showed in 2004 that "if traders are risk-neutral price takers with heterogenous beliefs, the price of a contract in a prediction market reveals nothing about the dispersion of traders' beliefs and partially identifies the central tendency of beliefs" [9]. The mean belief "lies in an interval whose midpoint is the equilibrium price" [9].

His worked example takes a contract priced at 0.75. That price "reveals that 75 percent of traders believed the probability of the event to be larger than 0.75", and the mean subjective probability "lay somewhere in the range (0.5625, 0.935)" [9].

Wolfers and Zitzewitz replied in 2006 that "for a broad class of models prediction market prices are usually close to the mean beliefs of traders" [10]. In their simulations with beliefs spread over 10 percentage points, the divergence "is typically within a percentage point, although the risk neutral model yields larger

differences", and "The divergence between prices and average beliefs is greatest for prices closest to $0 and $1" [10]. They conclude that prices "typically provide useful (albeit sometimes biased) estimates of average beliefs" [10].

Both papers put the weak point of the reading near the ends of the price range, where the favorite-longshot bias appears.

3. The favorite-longshot bias

Snowberg and Wolfers describe the bias as "the longstanding empirical regularity that betting odds provide biased estimates of the probability of a horse winning", in which "longshots are overbet, while favorites are underbet" [12]. Thaler and Ziemba reviewed racetrack and lottery betting in 1988, noting that "empirical research has uncovered several interesting anomalies" [11]. In "all 6,403,712 horse starts run in the United States between 1992 and 2001", the rate of return on horses at odds of 100/1 or more was about −61%, against −23% for betting at random and −5.5% for betting the favorite in every race [12]. The authors find evidence that "misperceptions of probability drive the favorite-longshot bias, as suggested by Prospect Theory" [12].

Prediction markets show the same pattern. Wolfers and Zitzewitz found in 2003 data that "the extremely unlikely (high and low) outcomes for the S&P 500 are relatively over-priced on Tradesports" against options prices [4]. Page and Clemen, using more than 500,000 Intrade transactions from markets that ran longer than a month, found markets "reasonably well calibrated when time to expiration is relatively short, but prices are significantly biased for events farther in the future" [13]. Long-term prices are "systematically biased towards 0.50", because time discounting makes holding a high-priced contract to a distant expiry unattractive [13].

Four recent studies measure the bias on current platforms:

Study	Data	Low-priced contracts	High-priced contracts
Bürgi, Deng and Whelan, September 2025 [14]	Kalshi, 2021 to April 2025, 313,972 contract prices	Win "far less often than required to break even"	Win more often
Becker, January 2026 [15]	Kalshi, 72.1 million trades	5-cent contracts won 4.18% of the time	95-cent contracts won 95.83% of the time
Clinton and Huang, June 2026 [28]	2024 political markets on Kalshi, Polymarket and PredictIt (IEM excluded)	Priced $0.10 to $0.20: average $0.149, 7.1% resolved Yes	Priced $0.80 to $0.90: average $0.847, 85.2% resolved Yes

Study	Data	Low-priced contracts	High-priced contracts
Cardozo and Rivero-Wildemauwe, September 2026 [16]	Polymarket, November 2022 to March 2026, 588 million trades	Purchases below 10 cents lost 19.3 cents per dollar	Bias present in aggregate, its size set by how contracts are grouped

Bürgi and colleagues found the pattern "across a wide range of different categories" of Kalshi markets, with "some evidence that the bias in prices is getting smaller over time" [14]. In their data, buyers of contracts under 10 cents lost more than 60% of their money, a measure of how far those prices sat above the win rate [14]. Becker still calls the overall curve evidence that "prediction markets are actually quite efficient and accurate, with the slight exception of the tails" [15].

The measured size depends on how contracts are counted. On Polymarket, the bias's "magnitude depends on how contracts are grouped": weighted equally, longshots lose 6.3 cents per dollar, and the sign of the result changes when related contracts are first grouped by parent event. The two-sided pattern was "surprisingly absent in Sports" [16].

4. Calibration across large samples

Nam Anh Le's single-author preprint, revised in August 2026, uses "353 million trades across 429,000 binary contracts on Kalshi and Polymarket", with trades to 31 December 2025 [17]. Le finds "persistent underconfidence in political markets, where prices compress toward 50%", a result that "replicates on Polymarket", with a mean slope of 1.45 there [17]. Across markets, the mean slope "rises from 0.99 (0–1 hour) to 1.32 (beyond one month)", while short-horizon weather markets move the other way [17]. Large political trades on Kalshi go with further compression, "a calibration-slope gap of roughly one-half", a result that "is not robust on Polymarket" [17]. Le concludes that "a price's meaning depends on what, when and how much is traded" [17].

Dickerson and colleagues published in PNAS on 19 October 2026 an analysis of "approximately one million resolved contracts and questions across four collective forecasting platforms, twelve domains, and time horizons ranging from hours to years" [18]. Its abstract reports that "events assigned a probability of p% occurred approximately p% of the time", while "fine-grained analyses revealed systematic tendencies toward over- or underestimation in highly consequential areas" [18].

The team's March 2026 preprint gives the detail: "1.014 million prediction contracts, collected between October 2015 and November 2025" from Polymarket, Kalshi, Manifold and Metaculus [19]. There, election forecasts "systematically underestimate outcomes", while forecasts on security and military matters, geopolitics and technology overestimate them [19]. When a market prices war at 75%, the preprint puts the likelihood closer to 63% [19]. Accuracy improved as markets neared expiry, while calibration improved only modestly [19].

Akey, Grégoire, Harvie and Martineau studied 729,133 Polymarket markets from November 2022 to March 2026, "covering over 2.4 million users and $67 billion in trading volume" [20]. They find prices "well-calibrated on aggregate", while "per-trade forecasting accuracy varies sharply across the cross-section" [20].

Brier.fyi, launched in 2025 "to help people compare and evaluate prediction markets fairly", tracks Kalshi, Manifold, Metaculus and Polymarket [22]. As viewed, one month before close "63% of markets were already within 30% of the correct resolution, representing a Brier score of 0.09 (n=84,357)" [21]. As of 14 October 2025, on 971 linked markets across 390 questions, "each platform's performance is similar on average", while "each platform outperforms the others in certain categories" [21]. Brier.fyi and the first author of the PNAS team's preprint both report grants from the Manifold Community Fund and EA Community Choice [19, 22].

5. Platform-run research on accuracy by horizon

Kalshi Research's working paper by Nicole Kagan and Rubens Baiocchi, as reported by Yellow.com and NEXTPredict on 21 August 2026, covers 2,243,741 resolved Kalshi markets in 11 categories from 2021 through mid-2026 [23, 24]. As reported, prices were "extremely well calibrated" in aggregate near resolution, with Brier scores, excluding combo contracts, falling from roughly 0.08 to 0.09 three months out to about 0.02 at close [23, 24], and naive accuracy rising from 88.3% to 97.2% [23]. In economics markets the score fell from 0.108 to 0.066 [24].

Kalshi split, close horizon, as reported	Brier score
Under $10,000 in event volume	0.0635 [23]
At least $200,000 in volume	0.0099 [23]
Fewer than 20 traders	0.0354 [23]
At least 1,000 traders	0.0089 [23]

The report calls trader count "an imperfect measure", and NEXTPredict adds that the volume split "does not establish causation" [23, 24]. Measured from when the event occurred, instead of the administrative close, the one-month Brier score for elections rose from 0.045 to 0.079 [24].

Polymarket's accuracy page, as viewed, scores how often the leading outcome matched the result at fixed points before resolution [25]:

Time before resolution	Leading outcome matched
4 hours	98.7%

Time before resolution	Leading outcome matched
1 day	96.6%
1 week	94.1%
1 month	90.4%

The page gives a Brier score of 0.0605 and a resolution composition of 21.7% Yes and 78.3% No [25].

That composition may explain part of the headline. Across all resolved markets on the page, a rule that always picked No would have matched 78.3% of them, though the accuracy shares come from 30,747 markets whose Yes-No mix the page does not give (derived from [25]). Kalshi has the same shape: about two-thirds of 313,972 recorded prices sat below 10 cents or above 90 cents, and "There are only 8,351 contracts with price between 50 and 59" [14]. In Le's Kalshi sample, 38.1% of contracts resolved Yes [17]. Naive accuracy counts a 51-cent winner and a 95-cent winner the same, so a book of lopsided contracts scores high on it [24].

6. The 2024 election study and the measurement dispute

On 1 December 2025, Joshua Clinton and TzuFeng Huang of Vanderbilt posted "Prediction Markets? The Accuracy and Efficiency of $2.4 Billion in the 2024 Presidential Election" [26]. It scored "more than 2,500 political prediction markets" on the IEM, Kalshi, PredictIt and Polymarket by log skill, using the closing prices of the 1,237 markets active on 4 November 2024 [26]. "While 93% of PredictIt markets correctly predicted outcomes better than chance, accuracy fell to 78% on Kalshi and 67% on Polymarket" [26]. It also found that "prices for identical contracts diverged across exchanges", and, on its PDF page 23, arbitrage opportunities on "62 out of the 65 days" in presidential winner contracts [26].

DL News reported the study on 5 December 2025 [27]. Kalshi's media relations lead said the methodology "completely misunderstands prediction markets, and shouldn't be taken seriously", and that "When measuring calibration, the true way to measure accuracy, Kalshi, is almost perfectly accurate" [27]. Clinton replied that "the point of our paper was not really about accuracy at all", and that similar markets were "not only consistently priced differently, but also that the changes in daily closing prices were largely unrelated" [27].

The version posted on 10 June 2026 calls itself "a dramatic update of prior versions" [28]. Its revision appendix lists the changes [28]:

- Each market is scored on "min(Nov 4, resolution_date−1)", because scoring on the day before administrative resolution "gave a mechanical 100% beat-coin-flip rate" for Kalshi.

- Polymarket prices come from "authoritative CLOB (Central Limit Order Book) daily prices" instead of an on-chain fill ratio, and resolutions from the platforms' APIs instead of hand coding.
- "Dollar-volume claims removed (i.e., the "$2.4 billion")", because Kalshi's API reports contracts traded and Polymarket's a liquidity or depth proxy.

Share beating a coin flip	Version of 1 December 2025 [26]	Version of 10 June 2026 [28]
Markets and window	1,237 scored (of more than 2,500 analyzed), final five weeks	2,170 scored, final nine weeks
Kalshi	78%	89.8% of 550
Polymarket	67%	85.3% of 1,587
PredictIt	93%	81.8% of 33

Across all 2,170 markets, 86.4% beat a coin flip, with a 95% interval of 84.9% to 87.7% [28]. Speech contracts did so 75.5% of the time and winner contracts 93.4% [28]. Polymarket's lower total "is driven almost entirely by speech contracts that comprise 45% of its total book", and "Across margin and other contracts, the exchanges perform almost identically" [28].

The markets "were also generally well-calibrated", "especially for contracts priced above $0.60", with "modest evidence of longshot bias" [28]. The expected calibration error was 0.032, and "the reliability component of the Brier score is 0.0022 (of a total Brier of 0.096)" [28]. The lowest price decile held 880 markets, 41% of the sample [28].

Composition carries the topline: 1,434 markets priced below $0.20 or above $0.80 beat a coin flip 96.2% of the time, 486 leaning markets 73.3% and 250 markets priced $0.40 to $0.60 55.2% [28]. Cross-platform price correlations for the winner contracts ran from about 0.77 to 0.96, and buying both main candidates' contracts for less than $1 across platforms, after trading costs, was possible "on nearly every day" of the 1 September to 4 November window [28]. The authors caution that past accuracy "neither guarantees future market performance nor implies that the price of any particular contract provides a definitive or fully efficient assessment of political reality" [28].

An April 2026 preprint, posted before the revision and built on the earlier evidence, asked "whether accurate forecasting is even the right criterion for a market that has become a public coordination device" [29].

7. Markets beyond elections

In a 2015 PNAS study, markets on 44 psychology studies in the Reproducibility Project "correctly predict the outcome of 71% of the replications (29 of 41 studies ...), which is significantly higher than 50%" (P = 0.012), while a simple average of a survey predicted 58% (23 of 40), not significantly different from 50% [30].

Wolfers and Zitzewitz's 2004 review describes internal corporate markets at HP and Siemens with "only a relatively small trading population (20-60 people)", and writes that "The predictive performance of even these very thin markets was quite striking" [4]. In the 2003 NFL season, real-money and play-money prices "yielded predictions that were approximately equally accurate", and both "outperformed all but a dozen of 3,000 people in an online contest" [4].

8. Markets as a baseline for LLM forecasters

Studies of LLM forecasters often score models against the market price on the same questions.

Study	Questions	Market price	Models
Prophet Arena, preprint revised December 2025 [31]	1,367 Kalshi events, 72,136 markets, to 11 October 2025	Brier 0.187 (±0.006)	Top of 23 models 0.184 (±0.006), four other highlighted models 0.189 to 0.219
Commercial forecasting system, technical report, November 2025 [32]	MarketLiquid: 1,610 questions from liquid prediction markets, resolving April to May 2025	Brier 0.1106	0.1258
The same report [32]	FB-Market: 76 ForecastBench questions drawn from prediction markets	Brier 0.0965	0.0753
FIFA World Cup 2026, preprint, July 2026 [33]	104 matches, three outcomes	0.4688	Four agents from 0.4705 to 0.4828

In Prophet Arena, "the market baseline lags behind several frontier LLMs when predictions are made far in advance", and near resolution markets take in news faster, "quickly surpassing LLMs in short-term accuracy" [31]. In the commercial system's report, on MarketLiquid, an ensemble with weights of 0.33 on the system and 0.67 on the market scored 0.106 in a leave-one-out estimate, against the market's 0.111 [32]. The World Cup preprint removed the bookmaker margin, a mean overround of 1.05, and found that "none beats the market's Brier score" [33]. Sethi and colleagues saw the same timing with a statistical model in 2020 [7].

9. What each study measured

Source	What was measured	Timing
Wolfers and Zitzewitz 2004 [4]	Absolute error of vote-share forecasts against the final Gallup poll	Final week
Berg, Nelson and Rietz 2008 [5]	Share of 964 poll comparisons where the market was closer	Up to and beyond 100 days out
Page and Clemen 2013 [13]	Calibration by time to expiry	Over each market's life
Bürgi, Deng and Whelan, Becker, Cardozo and Rivero-Wildemauwe [14, 15, 16]	Win rate or return by price band	Trade prices
Clinton and Huang, December 2025 [26]	Share beating a coin flip by log skill, price gaps across platforms	4 November 2024 close
Clinton and Huang, June 2026 [28]	The same, rescored, plus calibration and arbitrage	Last price before resolution, at most 4 November
Le 2026 [17]	Recalibration slope by domain, horizon and trade size	Under an hour to over a month
Dickerson and colleagues 2026 [18, 19]	Calibration across platforms and domains	Hours to years
Kalshi Research 2026, as reported [23, 24]	Brier score and naive accuracy by horizon, volume and traders	Three months to close
Polymarket accuracy page [25]	Share whose leading outcome won, and Brier score	Four hours to one month
LLM forecasting studies [31, 32, 33]	Brier score of models against the market price	Per study

Implications

For analysts reading a price as a probability. Across many contracts, the large studies support reading a price as an approximate probability [18, 20, 28]. For one contract the reading is looser: theory bounds the mean belief behind a price without fixing it [9, 10], and the 2024 study's authors warn against treating any single price as definitive [28]. The evidence points to adjustments by where the price sits and when. Cheap contracts have resolved Yes less often than their prices [14, 15, 28], long horizons pull prices toward 50% [13, 17], and political prices were compressed toward 50% in Le's data [17]. Thin markets scored worse in the platform-run Kalshi figures [23].

For teams using a market price as a baseline. The market price was hard to beat near resolution and easier far out, for an election model in 2020 and for LLM systems in 2025 [7, 31]. Combining a forecast with the market price improved on the market in two studies [7, 32]. The clock matters too: Kalshi's election scores worsened when anchored to the event [24], and in the 2024 study, scoring on the day before administrative resolution "gave a mechanical 100% beat-coin-flip rate" for Kalshi [28].

For readers of accuracy headlines. Three questions sort the figures. What was measured: closeness to polls, calibration, Brier score or the share called right [5, 18, 23, 25]. When: four hours before resolution gives a different number from one month [25]. On which contracts: a book in which 78.3% of markets resolve No, as across all of Polymarket's resolved markets, or in which 66% sit below $0.20 or above $0.80, scores high on the share called right [25, 28].

Rillor builds LLM forecasting agents for a client's own questions and data, and each forecast is stated before the outcome and compared with simple baselines, such as no change or the historical rate, and with the same pipeline run on randomly chosen dates.

Limits of this analysis

- The Kalshi Research working paper could not be opened. Its figures come from two news reports, and the platform's own research team wrote it [23, 24].
- Only the abstract of the PNAS article was read. Its counts, dates and the war example come from the team's March 2026 preprint [18, 19].
- Five sources were read from their published abstracts only [3, 6, 8, 11, 29].
- Many sources are preprints or working papers, including Le's single-author paper, read in versions not peer reviewed [14, 16, 17, 20, 26, 28, 31, 32, 33].
- Brier.fyi's and Polymarket's figures come from live pages as viewed on 8 and 14 November 2026 [21, 25].
- Le notes that both of his platforms are U.S.-centric, and his Polymarket comparison covers three of six domains [17]. Most election evidence here concerns U.S. presidential races.
- No study reviewed covers markets on the 3 November 2026 U.S. elections.
- Studies differ in measure, sample, platform and timing, so no figure here is compared across studies or averaged.
- Return figures in the longshot studies describe past data, and their size changes with how contracts are grouped [16].

Sources

1. Understanding Prediction Markets and Event Contracts. U.S. Commodity Futures Trading Commission, viewed 8 November 2026. https://www.cftc.gov/LearnandProtect/PredictionMarkets

2. Iowa Electronic Markets. University of Iowa, Henry B. Tippie College of Business, viewed 8 November 2026. https://iemweb.biz.uiowa.edu/

3. Prediction Markets. Wolfers and Zitzewitz, Journal of Economic Perspectives 18(2), 107-126. American Economic Association, Spring 2004, viewed 8 November 2026. https://www.aeaweb.org/articles?id=10.1257/0895330041371321

4. Prediction Markets. Wolfers and Zitzewitz, NBER Working Paper 10504. National Bureau of Economic Research, May 2004, viewed 8 November 2026. https://www.nber.org/papers/w10504

5. Prediction market accuracy in the long run. Berg, Nelson and Rietz, International Journal of Forecasting 24(2), 285-300. IDEAS/RePEc record, 2008, viewed 8 November 2026. https://ideas.repec.org/a/eee/intfor/v24y2008i2p285-300.html

6. Historical Presidential Betting Markets. Rhode and Strumpf, Journal of Economic Perspectives 18(2), 127-141. American Economic Association, Spring 2004, viewed 8 November 2026. https://www.aeaweb.org/articles?id=10.1257/0895330041371277

7. Models, Markets, and the Forecasting of Elections. Sethi, Seager, Cai, Benjamin and Morstatter, arXiv:2102.04936, version 4, 25 May 2021, viewed 8 November 2026. https://arxiv.org/abs/2102.04936

8. Are Betting Markets Better than Polling in Predicting Political Elections? Cutting, Hughes-Berheim, Johnson, Baroud and Goldstein, arXiv:2507.08921, version 1, 11 July 2025, viewed 8 November 2026. https://arxiv.org/abs/2507.08921

9. Interpreting the Predictions of Prediction Markets. Manski, NBER Working Paper 10359, published in Economics Letters 91(3), June 2006. National Bureau of Economic Research, March 2004, viewed 8 November 2026. https://www.nber.org/papers/w10359

10. Interpreting Prediction Market Prices as Probabilities. Wolfers and Zitzewitz, NBER Working Paper 12200. National Bureau of Economic Research, May 2006, viewed 8 November 2026. https://www.nber.org/papers/w12200

11. Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries. Thaler and Ziemba, Journal of Economic Perspectives 2(2), 161-174. American Economic Association, 1988, viewed 8 November 2026. https://www.aeaweb.org/articles?id=10.1257/jep.2.2.161

12. Explaining the Favorite-Longshot Bias: Is it Risk-Love or Misperceptions? Snowberg and Wolfers, NBER Working Paper 15923, published in the Journal of Political Economy 118(4), 2010. National Bureau of Economic Research, April 2010, viewed 8 November 2026. https://www.nber.org/papers/w15923

13. Do Prediction Markets Produce Well-Calibrated Probability Forecasts? Page and Clemen, The Economic Journal 123(568), 491-513, May 2013, author's copy, viewed 8 November 2026. https://people.duke.edu/~clemen/bio/Published%20Papers/45.PredictionMarkets-Page&Clemen-EJ-2013.pdf

14. Makers and Takers: The Economics of the Kalshi Prediction Market. Bürgi, Deng and Whelan, UCD Centre for Economic Research Working Paper WP25/19. University College Dublin, September 2025, viewed 8 November 2026. https://www.ucd.ie/economics/t4media/WP2025_19.pdf

15. The Microstructure of Wealth Transfer in Prediction Markets. Jonathan Becker, independent researcher's web paper, jbecker.dev, 18 January 2026, viewed 8 November 2026. https://www.jbecker.dev/research/prediction-market-microstructure

16. The Favorite-Longshot Bias in Prediction Markets: Evidence from Polymarket. Cardozo and Rivero-Wildemauwe, arXiv:2609.12878, version 1, 11 September 2026, viewed 8 November 2026. https://arxiv.org/abs/2609.12878

17. Decomposing Crowd Wisdom: Domain-Specific Calibration Dynamics in Prediction Markets. Nam Anh Le, arXiv:2602.19520, version 2, 4 August 2026, viewed 8 November 2026. https://arxiv.org/abs/2602.19520

18. Calibration, accuracy, and domain variation across one million collective forecasts from prediction platforms. Dickerson, Shinitzky, Peer, Ben-Shachar and Gilead, Proceedings of the National Academy of Sciences 123(43), 19 October 2026, abstract read from the DOI record, viewed 8 November 2026. https://doi.org/10.1073/pnas.2615627123

19. The knowability of the future: Evidence from one million collective forecasts. Dickerson, Shinitzky, Peer, Ben-Shachar and Gilead, PsyArXiv preprint, 12 March 2026, viewed 8 November 2026. https://osf.io/preprints/psyarxiv/g4ja9_v1

20. Who Wins and Who Loses in Prediction Markets? Evidence from Polymarket. Akey, Grégoire, Harvie and Martineau, working paper hosted by the Ontario Securities Commission under its research grant competition, 28 August 2026, viewed 8 November 2026. https://www.osc.ca/sites/default/files/2026-10/research-grant-competition-2026-who-wins-and-who-loses-in-prediction-markets.pdf

21. Brier.fyi. Project Themis, viewed 8 November 2026. https://brier.fyi/

22. About, Brier.fyi. Project Themis, viewed 8 November 2026. https://brier.fyi/about/

23. Kalshi Study Finds Prediction Markets Work Best When Traders Crowd In. Murtuza Merchant, Yellow.com, 21 August 2026, viewed 8 November 2026. https://yellow.com/news/kalshi-prediction-market-calibration

24. Large-scale Kalshi research examines calibration of event contracts. Erik Gibbs, NEXTPredict, 21 August 2026, viewed 8 November 2026. https://nextpredict.io/market-news/industry/large-scale-kalshi-research-examines-calibration-of-event-contracts/

25. How accurate is Polymarket? Polymarket, viewed 14 November 2026. https://polymarket.com/accuracy

26. Prediction Markets? The Accuracy and Efficiency of $2.4 Billion in the 2024 Presidential Election. Joshua D. Clinton and TzuFeng Huang, SocArXiv d5yx2, version 1, 1 December 2025, viewed 8 November 2026. https://osf.io/preprints/socarxiv/d5yx2_v1

27. Are Polymarket and Kalshi as reliable as they say? Not quite, study warns. Pedro Solimano, DL News, 5 December 2025, viewed 8 November 2026. https://www.dlnews.com/articles/markets/polymarket-kalshi-prediction-markets-not-so-reliable-says-study/

28. Prediction Markets? The Accuracy and Efficiency of Political Prediction Markets in the 2024 Presidential Election. Joshua D. Clinton and TzuFeng Huang, SocArXiv d5yx2, version 5, 10 June 2026, viewed 8 November 2026. https://osf.io/preprints/socarxiv/d5yx2_v5/

29. Price as Focal Point: Prediction Markets, Conditional Reflexivity, and the Politics of Common Knowledge. Nechepurenko, arXiv:2604.24147, version 1, 27 April 2026, viewed 8 November 2026. https://arxiv.org/abs/2604.24147

30. Using prediction markets to estimate the reproducibility of scientific research. Dreber, Pfeiffer and others, Proceedings of the National Academy of Sciences 112(50). PubMed Central, 9 November 2015, viewed 8 November 2026. https://pmc.ncbi.nlm.nih.gov/articles/PMC4687569/

31. LLM-as-a-Prophet: Understanding Predictive Intelligence with Prophet Arena. Yang, Mahns, Li, Gu, Wu and Xu, arXiv:2510.17638, version 2, 21 December 2025, viewed 8 November 2026. https://arxiv.org/abs/2510.17638

32. AIA Forecaster: Technical Report. Alur, Stadie, Kang and others, arXiv:2511.07678, version 1, 10 November 2025, viewed 8 November 2026. https://arxiv.org/abs/2511.07678

33. FIFA World Cup 2026 as a Contamination-Free Benchmark for LLM Forecasting Agents. Ding, Guo and Xu, arXiv:2607.17765, version 1, 20 July 2026, viewed 8 November 2026. https://arxiv.org/abs/2607.17765

About this report

Published by Rillor on 15 November 2026. Superintelligence (SI) tools assist research and drafting, and every figure cites its source. Rillor builds and secures SI systems, supplies the data they learn from and delivers the infrastructure they run on. We apply SI to financial and compute markets.

Cite as: Rillor. What research shows about prediction market accuracy. 15 November 2026. https://rillor.com/insights/prediction-market-accuracy-research

Notices

Rillor is not a registered investment adviser or commodity trading advisor and does not provide investment or trading advice. Research, forecasts, data and software described on this site are for research and engineering use. Nothing here is an offer or recommendation to buy or sell any security, commodity interest or digital asset. Past or simulated results do not indicate future results.

Superintelligence (SI) outputs can be wrong. Important decisions should include human review.

Work with Rillor

Send the question you want forecast, its horizon and the data you already hold. Contact us at rillor.com/contact?type=forecasting or contact@rillor.com. More on forecasting and market intelligence: rillor.com/markets/forecasting