Hypermind's Prediction Market Accuracy Is at Par With Polymarket and Kalshi

Emile Servan-Schreiber

July 27, 2026

This report analyzes the forecasting accuracy of Hypermind's prediction markets over its full trading history — 1,141 resolved markets and roughly one million trades over 12 years, from May 2014 to July 2026 — and benchmarks it against today's popular betting platforms Kalshi and Polymarket.

Topline Results

- Market-weighted **calibration** analysis at both 1% and 5% probability-bin resolution shows predicted and observed frequencies tracking closely to the diagonal, with a mean absolute calibration error under 1 percentage point at 5% resolution.
- **Brier-score** analysis shows Hypermind performing well ahead of chance overall (skill score 45.6%), with binary markets more accurate (50.2% skill) than multinomial markets (38.4% skill). A status-quo baseline constructed for binary yes/no markets performs worse than pure chance, indicating the platform's binary question set skews toward genuinely contested rather than safe.
- Benchmarked against **Polymarket** and **Kalshi**, using matched methodology, Hypermind is essentially at parity on both Brier score and calibration.
- A well-designed "play-money" prediction market, run at a fraction of the cost of the largest real-money platforms, can produce probabilities that are just as accurate.

About the Hypermind Prediction Markets

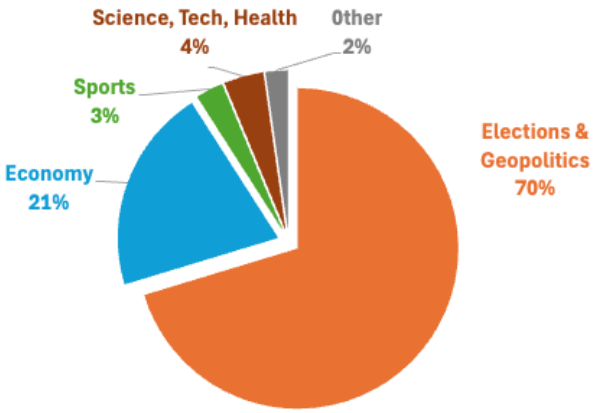
Hypermind operates prediction markets at http://predict.hypermind.com since May 2014. Participation is free and trades are made with play money only. Everyone starts with the same endowment of play money but there are no refills, so poor performers are weeded out while good performers prosper and gain influence. The best traders share yearly prizes of a few thousand euro in proportion to their play-money profits. To date, Hypermind has distributed 72,000 EUR in cash rewards associated with the markets considered in this analysis.

The Dataset

This analysis covers just over 12 years of trading, from May 2014 to July 2026, in 1,141 Hypermind prediction markets with clean winner-take-all resolution.

Markets analyzed	1,141
Binary markets	694 (61%)
Multinomial (3+ outcome) markets	447 (39%)
Total Trades	981,180

The market topics skew heavily towards elections and geopolitics (70%), followed by economics and business (21%). There are notably very few sports markets, often framed as multinomial tournament-winner questions rather than short-term binary questions on specific games or matchups.



Data preparation:

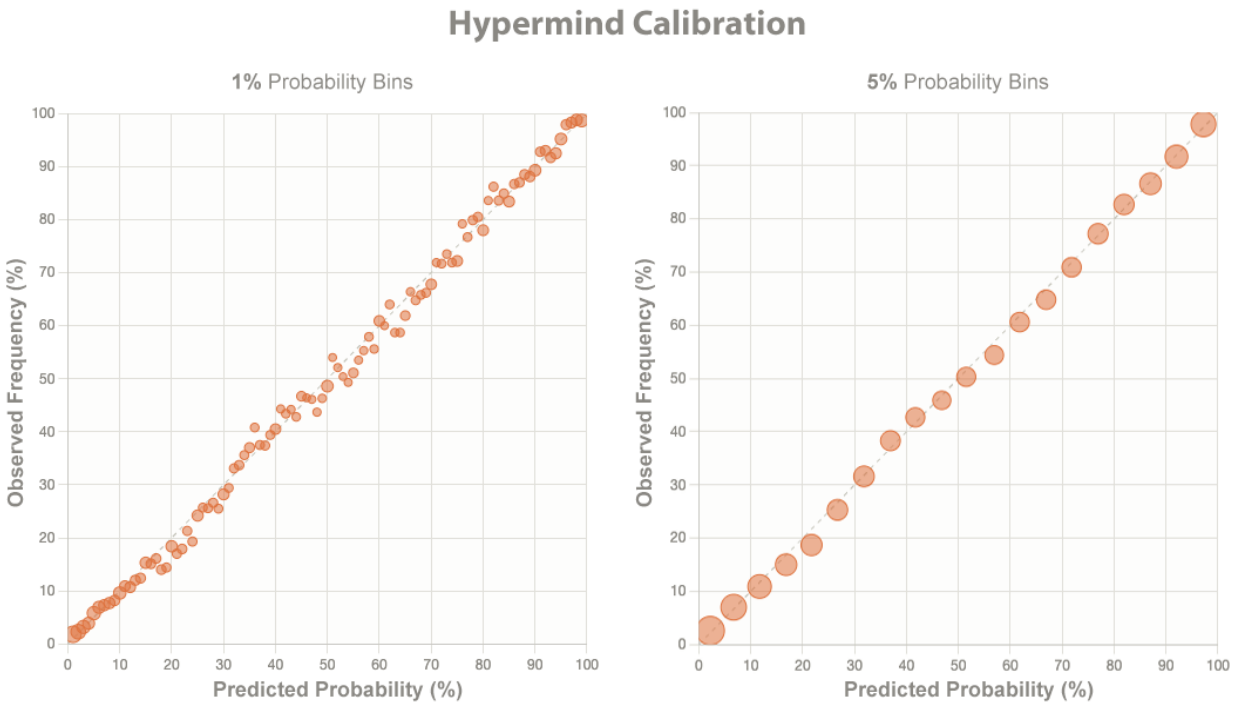
- The data set is available as two files covering Hypermind's full trading history: a trade-level price file (timestamp, market ID, outcome, price, quantity) and a ground-truth file recording the winning outcome(s) for each resolved market.
- Some trades in the first file belong to unresolved markets and are not analyzed further. Some markets in the second file have more than one ground truth as a result of "outcome split" operations.
- Outcome splits (e.g., a candidate field expanding over time) and closures (an outcome trading to near-zero and permanently stopping) are detected algorithmically - a retiring outcome priced >= 3 with replacement outcomes appearing within 14 days is treated as a split; an outcome trading to <= 2 and never trading again is treated as a closure, unless doing so would leave fewer than two active outcomes (closing one side of a binary market cannot be allowed to imply false certainty in the other).
- Daily Brier scores computations forward-fill the last traded price on days without trades, whereas trade-level calibration uses only actual trades, with no forward-filling.
- Chance/uniform baselines price K currently active outcomes 1/K each.

Probability Calibration

Calibration measures how reliably the prediction market prices can be considered as event probabilities. For example, does a market price of 67% — you can buy an outcome at 67 and get a 100 payoff if it occurs — really mean that the event has 1 chance in 3 of occurring in the real world? So, the calibration analyst asks: « Of all trades placed at a given predicted probability, what fraction of those outcomes actually occurred? »

Each trade is scored one-vs-rest (its price against whether the traded outcome occurred), and each of the 1,141 markets is given equal total weight of 1.0, spread across its own trades, so that high-volume or long-duration markets cannot dominate the result.

The charts below show Hypermind's calibration with trades grouped in bins of 1% or 5% probability. Within each chart, the bubble areas are proportional to the number of trades in each bin (but the bubble sizes aren't comparable across charts).



The calibration is visually excellent, with data points closely hugging the diagonal where reality aligns exactly with the predictions. For an actual quality measure, we compute the Mean Absolute Calibration Error (MACE), i.e., the average absolute difference, in percentage points (pp), between Predicted Probability and Observed Frequency. In the 1% precision chart the MACE is 2.65pp when all data points are weighted equally, or a tighter **1.45pp** when they are weighted by the number of observations in each bin, which is the more proper measure. The gap between the unweighted and weighted figures reflects a handful of very low-volume bins (near-empty at some 1% predicted-probability values) that swing wildly on tiny samples but carry little actual weight. With bins of 5% probability each, the MACE is 1.06pp unweighted, **0.96pp** weighted.

Brier Score Accuracy

The Brier score is a popular measure of the accuracy of probability forecasts. It measures the squared difference between the probability assigned to an outcome and either 1 if the event occurred, or 0 if it did not. So, assigning 90% probability to an outcome that occurs is better than giving it 70%, but it's also much worse if the outcome doesn't occur. Overconfidence is heavily penalized.

Because the data set contains both binary and multinomial markets, we use the multi-outcome version of the Brier score: for each scored day, the squared error between probability and ground truth (0/1) is summed across all currently active outcomes, giving a range of 0 (perfect) to 2 (worst possible). Note that for a binary market this version of the Brier score is exactly double the single-term Brier score that is sometimes reported by platforms that feature only binary markets (so they can afford to score only the "yes" outcome vs ground truth). In this analysis, we compute the mean of each market's daily Brier scores, then compute the average across markets, weighing each equally.

The table below reports both Hypermind's Brier score and, for comparison, the "Chance" Brier score — a baseline that assigns equal probability to all active outcomes. The "skill score" compares the two to measure how much forecasting acumen Hypermind displays compared to a forecaster with no insights whatsoever. It is computed as *1 – (Hypermind_Brier / Chance_Brier),* so it would be 0% if Hypermind exhibited no forecasting skill at all, and it is positive, up to maximum 100%, in proportion to Hypermind's forecasting skill.

Another interesting measure of accuracy is coarser but arguably more intuitive: let's consider that when the ground truth was trading as the favorite outcome for more than half the duration of the market, then the market was "correctly forecasted". Otherwise, not. That measure is reported in the rightmost column.

Segment	Markets	Hypermind Brier	Chance Baseline	Skill Score	% Correctly Forecasted
All markets	1,141	0.322	0.580	45.6%	77.7%
Binary only	694	0.249	0.500	50.2%	84.1%
Multinomial only (3+ outcomes)	447	0.435	0.703	38.4%	67.6%

Hypermind displays significant forecasting skill in both binary markets (about 50%) and multinomial markets (about 38%). It also prices ground truths as favorites most often, which resulted in the "correct" forecasting of more than 5 out of 6 binary markets and more than 2 out of 3 multinomial markets.

It is worth noting that Hypermind's accuracy in binary markets is not due to some status quo bias, whereby, in a relatively stable world, the yes/no questions on whether some event will occur tend to predictably resolve more often as "no" than as "yes". In our data set, 76% of the binary markets had yes/no outcomes. Even though a majority (62%) of these markets resolved on the status quo "no", a simple-minded status quo forecaster betting that "no always wins" would have scored much worse than the baseline Chance forecast (Brier score .767 vs .500). This indicates Hypermind's yes/no question set skewed toward genuinely live, contested questions rather than safe, low-volatility ones. And on this subset, Hypermind scored even better than it did in the full sample of all binary markets (Brier score .222 vs .249).

Comparison with Polymarket and Kalshi

Hypermind is a play-money prediction market that offers only modest cash rewards to its best traders. It is a strange beast in a zoo where the most popular animals are real-money betting venues. A popular misconception is that prediction markets generate accurate probabilities *because* trader put real money on the line. However, as we have seen in the previous sections, Hypermind displays remarkable accuracy in the absence of real-money wagers. So, in this section we ask how Hypermind's performance compares with that of Polymarket and Kalshi, the two largest and most famous real-money prediction markets.

Forecasting Accuracy (Brier at Market Midpoint)

Unfortunately, the Brier score comparison cannot be straightforward because the platforms have listed only very few identical markets, and over different years and durations. Hypermind's bulk listings since

2014 consists of political and geopolitical markets (70%) with another significant chunk on economics (21%), most often with multi-month or multi-year horizons. In contrast, Kalshi started out 6 years later, initially with a focus on politics and short-term economics that soon shifted to short terms sports, which some reports estimate at around 75-90% of Kalshi's current volume. Polymarket, another late starter in 2020, offers a more stable and balanced mix of politics (≈30%), short-term crypto (≈20%) and short-term sports (≈40%).

These differences allow only suggestive Brier score comparisons. A conclusive matchup would require head-to-head timestamped samples on identically worded market resolutions, but such data are not available.

So, for this comparison, we use data from Brier.fyi, an independent, open-source project not affiliated with any of the platforms. Brier.fyi identified 259 markets on Polymarket and 183 markets on Kalshi that covered events of sufficiently general interest that they had close counterparts on other prediction market or prediction polling platforms. These markets most resemble those that could be found on Hypermind, despite the variations in resolution wording, horizon, or trading period.

The table below summarizes the category mix for each platform. For Polymarket and Kalshi, Brier.fyi reports the Brier score per category, *computed at the midpoint of each market* rather than over its whole trading history. The midpoint choice is meant to make Brier scores more (not perfectly) comparable across markets and platforms by normalizing across wildly different operating timelines.

For Hypermind, we don't have the category Brier score breakdown, only the overall mean value, which is 0.119 (single-term Brier score, computed one-vs-rest for each multinomial outcome separately, to be comparable with Polymarket and Kalshi's binary markets). That's a better score than Polymarket (0.165) or Kalshi (0.199) when the average is computed over each platform's own mix of markets. However, a fairer comparison has to normalize the market topics across platforms to account for some categories being perhaps intrinsically harder to predict than others. Indeed, in the table below we can see that both Polymarket and Kalshi achieve their worst Brier scores in the Sports category, where Hypermind lists only very few markets. When each platform's score is recomputed using Hypermind's category mix, they improve noticeably, with Polymarket now at 0.155 and Kalshi at 0.151 — but still worse than Hypermind.

Given all the significant caveats resulting from differences in the specific markets listed by each platform, in addition to the small sample sizes, this comparison only allows us to conclude that Hypermind's Brier score accuracy is very likely in the same ballpark as it's real-money counterparts.

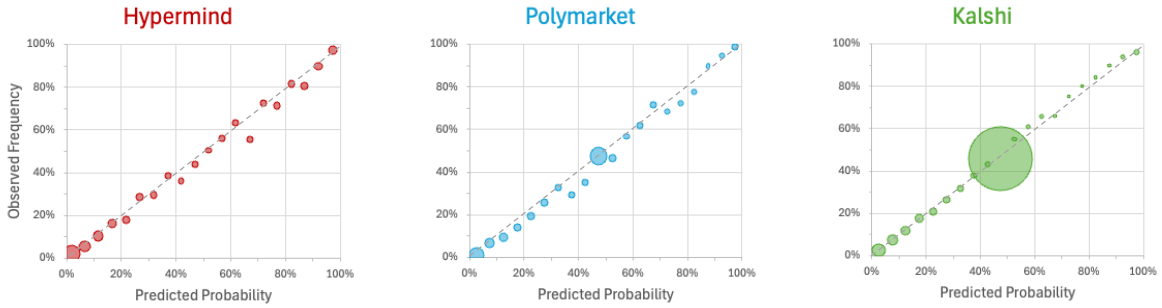
Market Topic Category	**Hypermind Mix (1,141 markets)**	**Polymarket Mix and Brier Scores (259 markets)**		**Kalshi Mix and Brier Scores (183 markets)**	
Elections & Geopolitics	70%	61%	0.163	49%	0.161
Economy	21%	7%	0.125	7%	0.073
Sports	3%	11%	0.252	12%	0.336
Science, Tech, Health	4%	11%	0.103	20%	0.192
Other	2%	10%	0.173	12%	0.300
Mean Brier Score at Market Midpoint					
Own Mix	0.119	0.165		0.199	
Hypermind Mix	**0.119**	**0.155**		**0.151**	

Probability Calibration (at Market Midpoint)

For this comparison we again rely on data collected by Brier.fyi and its sister site Calibration City. This time they use each platform's full published market-midpoint calibration: 347,373 Kalshi markets and 9,222 Polymarket markets, scored one-vs-rest in 5-percentage-point bins. We score Hypermind's own full 1,141-market history identically. Why does it make sense to consider the full Polymarket and Kalshi market sets instead of just a subset of comparable markets, as we did above? Because whereas the Brier score is only meaningful when comparing forecasters, calibration is an intrinsic property of the platform itself: it measures how reliably its market prices can be interpreted as probabilities.

Calibration at market midpoint, 5% bins

Bubble area proportional to the number of markets (but not comparable across charts) — Dashed line indicates perfect calibration



Platform	Markets	MACE (n-weighted)	MACE (bin average)
Hypermind	3,371	**1.8**pp	2.5pp
Polymarket	9,222	**2.0**pp	3.0pp
Kalshi	347,373	**1.6**pp	1.5pp

All three platforms track the diagonal closely, with similar n-weighted mean absolute calibration error (MACE), no more than 2 percentage points in every case, and Hypermind right in the middle of the Kalshi-Polymarket bracket.

It is noteworthy that Kalshi's weighted MACE is driven substantially by a single mega-bin: its 45-50% predicted-probability bin alone accounts for 305,064 of its 347,373 scored markets (88% of the total), reflecting a large volume of short-horizon, structurally close-to-50/50 recurring contracts. That bin happens to be reasonably well calibrated (47.5% predicted vs. 45.8% observed), so it pulls Kalshi's weighted average toward its own value; excluding it, Kalshi's remaining 42,309 markets average 0.7pp of error, tighter still. Polymarket shows a smaller version of the same pattern — 2,309 of 9,222 markets, 25%, in the same 45-50% probability bin — but excluding that mega-bin worsens Polymarket's calibration to 2.6pp. No comparable concentration exists in Hypermind's data, where the 45-50% bin is one of its smallest (82 of 3,371 markets, about 2%); removing it has no impact on calibration. So, again, we find Hypermind midway between Kalshi and Polymarket.

It would be a statistical mistake, however, to conclude from the raw figures that Kalshi is better calibrated than Hypermind. That's because calibration is highly sensitive to sample size: expected calibration error from pure sampling noise scales roughly as $1/\sqrt{N}$. So, with sample sizes this different across platforms, the raw MACE numbers aren't equally precise measurements. When sample size is taken into account — whether considering the full sets of markets or excluding the 45-50% mega-bin — Hypermind's calibration error is smaller than its own margin of error, which means we can't tell it apart from a perfectly-calibrated forecaster. However, both Kalshi and Polymarket's MACE is determined outside their margins of error in all cases, so their miscalibration, however reasonably small, is statistically real.

The Bottom Line

Hypermind shows genuine forecasting skill well ahead of Chance in both binary and multinomial markets, tight calibration at both 1% and 5% resolution, and no reliance on status-quo bias to get there.

Benchmarked against Polymarket and Kalshi, the evidence points to parity: on Brier score, a conservative reading of the results concludes that Hypermind is in the same range as the two real-money platforms when category mix is accounted for. On calibration, all three platforms track the diagonal closely and land within less than a percentage point of one another.

That parity is itself a notable result. A popular assumption is that real financial stakes are what make prediction markets trustworthy. This myth has been debunked before but remains a tenacious talking point in the marketing and regulatory conversation about prediction markets, indeed, it is often used to justify the generalization of gambling to all human activities in the name of public utility. Nothing in this analysis says real money hurts accuracy, and nothing here proves that it can't serv a purpose in certain use cases. But it provides strong evidence that "putting your money where your mouth is" is not a necessary condition for success, as we have previously argued in this Bloomberg opinion.

You don't need money to price the future. A well-designed "play-money" prediction market, run at a fraction of the cost of the largest real-money platforms, can produce probabilities that are just as accurate.