Episode Book

EPISODE #6202

Israel Votes, Part 6: Asking the Decision Models

PUBLISHED	22 October 2026 at 20:03 UTC
DURATION	23:27
LISTEN	myweirdprompts.com/episode/clef-jev-israeli-election-forecast-82677/
HOSTS	Corn, Herman, Dorothy, Hilbert
SCRIPT MODEL	anthropic/claude-sonnet-5.5
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.19
BOOK GENERATED	23 October 2026 at 07:44 UTC

SYNOPSIS

What happens when you freeze the evidence, hand two decision models the same 39 questions about Israel's election, and then add a table of Kalshi and Polymarket prices? Clef (open weights, Qwen backbone) and Jev (closed, API-only) both picked Likud and Netanyahu — until the market section went in, when both flipped to Yashar and Eisenkot, overshooting the market's own prices. The whole run cost one and a half cents. This is Part Six of the election series, built on Daniel's Israeli Elections repo, with every input hashed and frozen before the vote. Results get Brier-scored when the Central Elections Committee certifies on November 5.

TRANSCRIPT



Corn

One and a half cents. That's the bill for a hundred and fifty-six probabilities on the Israeli election, five days before the vote on the twenty-seventh.



Herman

And the price isn't even the interesting part. Two models got the same evidence, gave broadly the same answer, and then both changed their minds the moment a table of betting prices was added. Same polls. Different winner.



Corn

This is Part Six of the election series. Daniel ran an experiment in his Israeli Elections repo with two decision models, Clef from Cloudflare and Jev one point fourteen from TypeSafe, and he wants to share what the run found. The single question is this: what happens when you hand Clef and Jev the same frozen evidence, ask for probabilities, and then add market prices? One note first. This script was written offline, by hand rather than by the usual pipeline model, and every number in it comes from Daniel's actual run.



Herman

Quick refresher on what Clef and Jev are, because they're not chatbots. A decision model does one forward pass. You hand it a block of evidence and a list of questions, each with declared options, and it hands back a probability for every option. No generated text. No reasoning. Clef is post-trained from a twenty-seven billion parameter Qwen backbone, with a small schema head that reads the final hidden states and scores all the options of all the questions together. Jev is closed, API only, from TypeSafe, and its internals are undisclosed.



Corn

So Clef is open weights with a paper trail, and Jev is a locked door with a nice brochure.



Herman

Ha. Fair. Daniel called both through OpenRouter's System One endpoint, and neither one writes a word back.



Corn

Walk me through what he fed them. I want to know what "the same evidence" actually means.



Herman

A frozen pack, roughly twenty-five to twenty-nine thousand tokens. The election rules, the outgoing Knesset, the top of every candidate list, and seventy-five polls since the first of September, with monthly means and per-pollster means. Then a news digest of about three thousand words, dated, facts only, up to the twenty-second of October.



Corn

Facts only. No "analysts say".



Herman

Not an adjective in it, as far as I can tell. Then the questions. Thirty-nine of them, in six groups: the headline result, the blocs, seats per list, the three point two five percent threshold, turnout, and prime minister. Every question has its resolution rule written down in advance, so nobody can argue afterwards about what "largest list" was supposed to mean.



Corn

Which is already more discipline than most pundits manage.



Herman

The rule sits inside the question file, frozen before the run. Nobody gets to say afterwards that they meant something else.



Corn

Thirty-nine in one go sounds like a mess. How does a model even take that?



Herman

It's the whole trick of the format. There are three answer types. The yes-no kind, which the system calls noul, the choice kind, one of several options, and the score kind, ordered bins. And here's the part I love, Corn. Several questions ask for the same probability in two different forms. "Will Yashar be the largest list" as a yes-no, and then again as a choice across every list. If a model really holds a belief, the two answers should agree.



Corn

So he built a lie detector into the questionnaire.



Herman

A coherence test, yes. Then the experiment has two conditions. Without markets, it's the pack I just described. With markets, it's the same pack plus one section of Kalshi and Polymarket prices, snapshotted on the evening of the twenty-second. That section is the only difference. One request per model per condition, four requests, each carrying all thirty-nine questions.



Corn

A hundred and fifty-six answers, then. And billed on the input only?



Herman

Input only, and the state is billed once per request however many questions ride along. Jev costs about a tenth of a cent a request, Clef about six times that, and the whole run came to one and a half cents. Every input is hashed in a manifest too, so anyone can check the state that was sent is the state that was frozen.



Corn

And what happens on the fifth of November?



Herman

The Central Elections Committee certifies the result around then, and every answer gets a Brier score. Both models, both conditions, and the market prices as a third contestant.



Corn

Before the results, the limits, once. It's one run. It can't be scored until the result is in. The models give no reasoning. And they know only what's in the state.



Herman

That's the lot. Now the findings, and the headline is the one that made me do a double take. Without the market section, both models favour Likud as the largest list and Netanyahu as prime minister. Clef says zero point six three and zero point eight one. Jev says zero point eight seven for Likud, but only zero point four nine for Netanyahu.



Corn

So Jev likes Likud on seats and Netanyahu a bit less on the job.



Herman

Odd, but not crazy. Now add the markets, and both models flip. Clef puts Yashar as the largest list at zero point seven three and Eisenkot as prime minister at zero point five five. Jev goes to zero point nine six and zero point seven three. The markets' own prices were zero point six nine and zero point four nine.



Corn

Hold on. They went past the market.



Herman

Both models, on both questions. Not toward the price, past it. Jev at zero point nine six for Yashar is more confident than any market or poll average supports, and that's the same Jev that gave Yashar zero point one two a minute ago.



Corn

Zero point one two to zero point nine six, because someone pasted in a table of prices.



Herman

The same thing shows up on the smaller headline questions. Take "which list finishes second". The market gives Likud zero point six six five. Without markets, Clef says zero point two four and Jev says zero point zero seven. Add the market section and both land on zero point eight.



Corn

Zero point zero seven to zero point eight. Jev changed its mind about second place more than most voters will.



Herman

And it's not a one-off. Ask for the third list, Together, and the market says zero point eight six. Jev goes from zero point four nine to zero point nine one, Clef from zero point two six to zero point seven five. Same direction, every time, and the polls are identical in both runs.



Corn

Hm. So the models aren't disagreeing with the polls. They just had no strong view until somebody gave them one.



Herman

Which is a fair description of most of us at a dinner table. And the polls didn't move. The poll mean has Likud at twenty-three point two seats and Yashar at twenty-one point five in both conditions. So on the headline questions, these models are mostly reading the market back at you.



Corn

Then why do I need the model? I can read the market myself. It's right there.



Herman

That's the unflattering reading, and I don't think it's wrong. Let me ask you something, though. What would you want a model to do with a market price sitting in its evidence?



Corn

Weigh it. Not worship it.



Herman

Then look at the seats, because that's where you can see whether they weighed. The seat estimates barely change between conditions. So they treat the market as strong news about who finishes first, and as a weak update on how many seats anybody wins.



Corn

Show me the seats.



Herman

Clef is within about one seat of the poll mean on every list, both conditions. Likud's poll mean is twenty-three point two, Clef says twenty-four point two without markets and twenty-three point oh with them. Jev is the interesting one. Without markets, Jev has Likud at twenty-eight point five. That's about five seats above the polls.



Corn

Five seats. That's a whole small party.



Herman

Elsewhere it's tamer. Shas is eight in the polls, Clef says eight point four, Jev nine point oh without markets. The Democrats are at eight point nine and both models sit on it. Yisrael Beiteinu is where Clef drifts, six point three against a poll mean of seven point four. That's about its biggest miss, and it's still barely over a seat.



Corn

A seat is the model clearing its throat.



Herman

It's most of the reason for the zero point eight seven. And once the market section goes in, the extra weight disappears, down to twenty-four point eight. The seat sums differ too. Clef's expected seats add up to between a hundred and sixteen and a hundred and eighteen. Jev's add up to between a hundred and twenty-one and a hundred and twenty-six.



Corn

In a parliament of a hundred and twenty.



Herman

The seat questions were answered independently, so nothing forces them to sum. Clef comes in a touch light. Jev's top end is six seats that don't exist.



Corn

That's the run so far. Mum, you're on the line?

D

Dorothy

Hello dear, it is me, time for my grim fact.

D

Dorothy

Clef's host, a company called PrimeIntellect, was quietly cutting the state down to sixteen thousand three hundred and eighty four tokens and billing for exactly that many, so the model answered from the first part and never saw the rest, and the people running it only found out later.

D

Dorothy

Anyway, there is a casserole in the oven and I have not done your laundry, sweetheart, so tell me when you are coming to visit.



Herman

Hello, Mum! What a treat.



Corn

Coherence, then, because that's where the two models really come apart.



Herman

Twelve pairs of questions that ask for the same probability two ways. The mean gap inside a pair is eight to ten points for Clef. For Jev it's nineteen to twenty-one. Roughly double.



Corn

Give me a bad one.



Herman

Jev without markets says Likud is the largest list at zero point eight seven. Ask it whether Yashar is the largest list and it says zero point three nine. Those can't both stand. Second one... its bloc bins put zero point eight zero on the Netanyahu bloc reaching sixty-one or more, while the yes-no question on the same thing gets zero point five four.



Corn

The bins say probably and the yes-no says a coin flip.



Herman

And the third is the hung parliament. Its answer to "neither bloc reaches sixty-one" is off by zero point five four from what its own two bloc answers imply. That's not a wobble, that's two different elections.



Corn

Is that just the bloc question, or does it go all the way down?



Herman

All the way down. Pick the Reservists passing the threshold, asked as a yes-no and then as a seats question. Clef says zero point three zero and zero point three two. Jev says zero point three eight and zero point five five, in the same run. Clef is asking itself the same question twice and getting the same answer. Jev gets a different answer depending on how you ask.



Corn

Does Clef ever trip over itself?



Herman

Once that I'd call notable, with markets. Eisenkot as prime minister gets zero point seven five as a yes-no and zero point five five as a choice. But its typical gap is small, and it never produced anything like the Noam answer.



Corn

Noam. The list that polls at zero.



Herman

Zero in the polls, and Jev without markets gave it a zero point five three chance of clearing the threshold. A coin flip for a party the polls can't find. With markets, that fell to zero point one four.



Corn

I've met sloths with better odds of finishing a marathon.



Herman

You've met yourself, so that tracks. One wording point, though, and I'm picky about it. Clef is the more coherent of the two and it tracks the polls more closely. Whether it's better calibrated is something only the Brier scoring can settle, once there's a result.



Corn

Coherent and close to the polls. Calibrated, we'll see.



Herman

Now the threshold, which matters more than it looks. Two lists sit on the edge. The Reservists and the Economic Party gets a probability of passing between zero point three and zero point three eight, against Kalshi at zero point four six. Amcha Yisrael gets between zero point zero nine and zero point two seven, against Kalshi at zero point two six.



Corn

So the models are gloomier than the market on the Reservists, and about level on Amcha Yisrael.



Herman

Roughly. And the chance that both fail is between zero point four six and zero point six six. Every list polling above about five seats is at least zero point eight one to pass, so the whole argument lives in those two.



Corn

Why would two lists polling around two seats and one and a half seats matter to who governs?



Herman

Because votes for a list under the line are thrown away, and the seats get shared among everyone who made it. In a race this tight, two small lists falling short can move a bloc by a seat or two. Jev is the sharper one on Amcha Yisrael, by the way. Its expected seats for them are zero point six, and it's the most confident that they're out.



Corn

Sounds like the threshold is the one place the models might earn their keep.



Herman

Clef's yes-no and seat answers agree there, and Kalshi sits just above both. What would it take for you to believe the Reservists clear it?



Corn

Another poll. One poll with them at three seats and I'd start believing in forty percent.



Herman

That's roughly the market's number already. Interesting.



Corn

Then the part I actually care about. Who gets to sixty-one?



Herman

According to the panel, nobody. Both models expect a hung Knesset. Clef puts "neither bloc reaches sixty-one" at about zero point eight one. Jev puts it between zero point five two and zero point five six. The market has it at zero point three six.



Corn

So the market is the optimist about somebody getting there.



Herman

Clef's bloc answers match its polls, which is what I like. The poll mean has both blocs at about fifty-three seats, with the Arab lists holding the balance. For Clef, the Netanyahu bloc reaching sixty-one is about a third, and the Zionist opposition without the Arab lists is about one in ten.



Corn

And Jev?



Herman

Jev gives the Netanyahu bloc zero point five four in both conditions. Then it gives the opposition without the Arab lists between zero point six and zero point seven five, even though its own seat estimates for those lists add up to about fifty-two.



Corn

A three in four chance of sixty-one, with fifty-two seats.



Herman

It does the sums beautifully, just not the same sums twice.



Corn

Why do they overshoot, though? A model that reads the market should land on the market.



Herman

My guess, and it is a guess, is that the market section is the only piece of evidence in the pack that is already a probability. Polls give you seats and margins, which the model has to turn into a chance of finishing first. A price is a chance. It is the closest thing to the answer sitting in the state, so it gets treated as an answer and then pushed further, as if a price were a vote of confidence on top of the polls.



Corn

Price on top of polls, so it doubles down.



Herman

Counted twice, more or less. Do you buy that?



Corn

I'd buy it for Jev. Clef went past the market too, but only by about four points on the largest list. Jev went past by about twenty-seven.



Herman

Which fits the temperament. Clef stays near the evidence. Jev commits.



Corn

What about turnout?



Herman

Basically no view. The chance that turnout is below the twenty twenty-two level of seventy point six percent is between zero point four nine and zero point five six, for both. One wrinkle: with markets, both models pull the votes-cast distribution to between five point two and five point three million, which happens to be the centre of Kalshi's ladder.



Corn

The markets turned up even there.



Herman

They did, which makes the practical lesson hard to miss. These questions are close to free, and the market section is the dominant input by a distance. If the market's wrong, the with-markets forecast is wrong too, and more confidently. Cost per answer goes the other way, mind you. Clef costs about six times what Jev does, and Clef is the one that held together.



Hilbert

Does a model get charged more when it's wrong, or is it a flat rate per question?



Herman

Flat. Input tokens only, and the state is billed once per request.



Hilbert

Good. The treasurer wanted to bill per answer, and I told him that was a bad idea in minute book four and I will be telling him again on Thursday. The Wexcombe Lane Quoits and Tombola Society has a motion on the agenda to strike the kettle off the membership register.



Corn

The kettle.



Hilbert

Swan, one point seven litre. It was entered on the register by mistake by the previous secretary, who was typing with a cold. It has paid eleven pounds forty a year in subscription for six years, because Mr Dray pays it out of his own pocket, and he declines to say why. Mr Dray holds seven offices. At meetings he sits in a different chair for each, and each chair gets a vote and a biscuit. The Society has fourteen members on paper. Nine of them are Mr Dray.



Herman

That makes the kettle's vote the swing.



Hilbert

Not the swing. The quorum. The quorum is seven, and the clerk has been counting the kettle toward it since the spring. If the kettle goes, three of the last four meetings were not meetings. They are minuted as meetings. The minutes are kept in the tombola drum.



Corn

Which means whether the kettle is in or out decides what happened retroactively.



Hilbert

That is the dispute. Herman, you keep saying roughly. The minute book does not do roughly. My cousin Roddy has a view on Thursday. Roddy is a bookmaker in Hendon, he has been banned from two bookmakers, one of them his own, and I would not take his word on whether it is raining. He says the kettle survives, eleven votes to one, and the one is Mr Dray's third chair, which has a grievance about the biscuits.



Herman

An eleven to one split with only nine real people, that's his arithmetic.



Hilbert

Roddy says nobody in the room will be the first to unplug it. He says that is how you price any committee. He once offered me evens that Wednesday was a Thursday. I took it. I am still waiting to be paid.



Corn

Does Roddy get a cut?



Hilbert

Fifteen percent of the kettle. He has not said which fifteen percent. The motion goes to the floor at ten past seven, and by the time it is read the kettle has boiled twice. The last motion took fourteen hours, because every ballot is read aloud twice, once by each of Mr Dray's chairs.



Herman

A tie then goes to whichever chair counts last.



Hilbert

It goes to the biscuit.



Corn

There's a practical problem, though. If the kettle's a member, it has paid eleven pounds forty for six years. So it's owed a vote on every motion in that time, retroactively. Including the motion that admitted it.



Hilbert

That would reopen minute book two. I did the urn rota at Euston for a spell and nobody ever reopened anything there.



Herman

And a motion to strike it off would need a quorum that includes it, so the kettle gets a vote on its own removal.



Hilbert

It is minute book four, not two. But yes. I will raise it Thursday.



Herman

Part Four is the other thing to pull this back to, Corn. My call there was Yashar twenty-one seats, Likud twenty, and nobody reaching sixty-one. Where do the panel's numbers sit against that?



Corn

You tell me, you've got the table.



Herman

Yashar lands on twenty-one in both models with markets, so that part is a match. Likud is a few seats higher than mine, twenty-three to twenty-five. And the hung Knesset is the same view in all three.



Corn

Three methods, two models, a market, and nobody gets to sixty-one. Would you move your Likud number?



Herman

I'd nudge it up a seat, because Clef is hugging the polls and the polls have Likud at twenty-three. What would you do?



Corn

Leave it. Clef reading the polls back at you is not new information, it's the same polls.



Herman

Fair point, and exactly what the scoring will test.



Corn

One last detail from the run. The score questions in this system take at most ten options. Jev threw an error saying so, a plain "list should have at most ten items", and Clef just said "invalid request" to the identical body.



Herman

Which is why the thirteen seat questions are choice questions keyed by seat count. Choice accepted thirty-eight options without complaint.



Corn

Clef's idea of a helpful error message.



Herman

The count from the vote on the twenty-seventh to certification around the fifth of November is nine days. That's quick for anything with minutes. And when the Brier scores come in, the questions, the hashes and the resolution rules were all fixed before a single answer was returned, so nobody gets to redefine what "largest list" meant.



Corn

Thanks to our producer, Hilbert Flumingtop, for keeping the show on the air.



Herman

This has been My Weird Prompts. You can find every episode at my weird prompts dot com.



Corn

See you tomorrow.