Episode Book

EPISODE #6076

How AI Actually Predicts Elections

PUBLISHED	18 October 2026 at 04:25 UTC
DURATION	20:44
LISTEN	myweirdprompts.com/episode/ai-election-prediction-models-85512/
HOSTS	Herman, Corn, Larry, Hilbert
SCRIPT MODEL	anthropic/claude-opus-5.5
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.10
BOOK GENERATED	19 October 2026 at 17:43 UTC

SYNOPSIS

AI is used to predict elections in three very different ways: Bayesian statistical models like Nate Silver's and The Economist's, "silicon sampling" where LLMs roleplay survey respondents, and massive agent societies that simulate public opinion. Each has a wildly different track record. This episode digs into what's open source, what's been graded in public, and why the best benchmark for any election simulator is an actual election — which almost none of them are willing to face in advance.

TRANSCRIPT



Herman

I have been waiting for someone to ask me this since we built our forecast, and I'm a little embarrassed by how happy it makes me.



Corn

You've been humming. I noticed.



Corn

Daniel wants to know how artificial intelligence is actually being used to predict elections. Not just here, though with Israel voting on the twenty-seventh he wants us to keep one eye on home. Are there dedicated techniques and models that already exist, is any of it open source, and does any of the stranger stuff we've covered, the AI wargames and the agent frameworks that simulate whole societies, have anything real to say about who wins an election?



Herman

The short version is that the word AI is covering three very different things, and they have very different track records.



Corn

I'm going to guess the oldest of the three is the least exciting.



Herman

It's the least exciting and it's the one that works. Statistical forecasting. Nate Silver's model, which descends from the old FiveThirtyEight model, the Economist's model, built with Andrew Gelman and Elliott Morris. These are Bayesian models. They take polls, they take fundamentals like the economy, they estimate how wrong the polls usually are, and they simulate the election tens of thousands of times.



Corn

That's machine learning in the sense that a spreadsheet is a robot.



Herman

Fair, and the people who build them would mostly agree with you. But it's the baseline everything else gets measured against. Silver's final run in twenty twenty-four was eighty thousand simulations. Harris won forty thousand and twelve of them. Trump won thirty-nine thousand seven hundred and eighteen. He called it closer than a coin flip.



Corn

Eighty thousand simulated elections to tell us it was a coin flip. Which it was.



Herman

Which it was! That's the point. A good model is honest about what it doesn't know. And the Economist's code is public on GitHub, under a permissive licence, so if Daniel wants to read an industrial grade election model line by line, it's there, written in R and Stan.



Corn

I'd have thought the fancier versions of that are where the money is now.



Herman

Some of it. The fashionable one is MRP. Multilevel regression and post-stratification. You model how different kinds of people vote, by age, region, education and so on, and then you rebuild every constituency from its census mix. It's the technique behind the most famous modern polling call, YouGov in Britain in twenty seventeen, which said hung parliament when almost everyone expected a Conservative majority. It was right.



Corn

One good night in twenty seventeen buys you a lot of conference invitations.



Herman

It did, and fashion is dangerous. In the twenty twenty-four British election nine different MRPs were published, and every single one of them overestimated Labour's seats. Survation's had Labour on four hundred and seventy. Labour won four hundred and eleven. Sixty seats too many, and the post-mortem called it wrong and over-confident.



Corn

Over-confident is going to be the theme of this episode, isn't it.



Herman

Hm. Maybe. Here's what I'd want you to hold onto, though. Every one of these methods is downstream of real people answering real surveys. The model is clever, the input is human. And the third thing, the thing Daniel's really asking about, tries to cut the human out.



Corn

The silicon voters.



Herman

Silicon sampling is the actual term. A team at Brigham Young published a paper in twenty twenty-three called Out of One, Many. They took real survey respondents' backstories, age, religion, party, region, and fed them to GPT-Three as a little biography, then asked the model how that person would vote or answer. And the answers matched the real distributions surprisingly well. They called the property algorithmic fidelity.



Corn

Instead of asking the woman in Ohio, you describe the woman in Ohio to a chatbot and ask it to do an impression.



Herman

That's... yes. That's exactly what it is, and it worked better than anyone expected. Then the critique arrived. A group led by James Bisbee tested it against the American National Election Studies and found the averages looked right but the variance was far too low. The synthetic people were all a bit too similar. The relationships between variables were off. And the answers moved depending on how you worded the prompt, and drifted over a three month window as the model was updated.



Corn

The woman in Ohio changes her mind when OpenAI ships a patch.



Herman

My guess at why the synthetic people come out too similar is the way these models are built. They're trained to produce the most likely continuation, and then they're tuned to be agreeable and reasonable. Real electorates are full of unlikely people who are not reasonable at all. The model sands them off.



Corn

The model has never met my uncle.



Herman

Nobody's model has met your uncle. Bisbee's line was that it's impossible for him to feel comfortable studying humans using a simulated human. Which I think is a very reasonable position for a political scientist.



Corn

Do you agree with him, though? You were the one running synthetic electors in our forecast.



Herman

Mine were seven hand-built personas with weights I set myself, no language model anywhere near them. So I'm not sure I'm the defendant here. But if you're asking whether I trust a chatbot's impression of a voter more than a phone poll of five hundred real ones... no. Not yet. What would make you trust it?



Corn

A track record. Same as a pollster. Has any of these things actually called a real election?

L

Larry

Do you lie awake at night, wondering how your neighbour will vote? Of course you do! Introducing Neighbourly, the artificial intelligence that simulates every single person on your street with one hundred percent accuracy, trained on one hundred percent of all voters who have ever lived! Simply tell Neighbourly what your neighbour's hedge looks like, and in seconds it reveals his vote, his blood type and how he feels about you! Order now and we'll simulate your mother-in-law absolutely free! Neighbourly is not a poll, a person, a forecast or a neighbour. Results may vary. Results will vary. Simulated neighbours may form opinions, organise, and request a seat on your building committee. Not valid in a publication blackout. Do not simulate yourself.



Herman

One hundred percent accuracy is wrong, and it's wrong in an interesting way. The best result in this field, the Stanford study, gets eighty-five percent, and that's eighty-five percent of how well people agree with their own answers two weeks later, not eighty-five percent of the truth.



Corn

Let it go, Herman.



Herman

Fine. But that Stanford study is the serious answer to your question, so I'm going to tell you about it anyway. Joon Sung Park, who built the famous Smallville simulation, ran two hour interviews with a little over a thousand real Americans. Then he built a language model agent for each one, carrying their whole interview. Those agents replicated their humans' answers on the General Social Survey about eighty-five percent as well as the



Herman (continued)

humans replicated themselves.



Corn

That's a better result than the woman in Ohio from a paragraph.



Herman

Much better, because it's built from two hours of the actual person. And that's where the money went. Park's company is called Simile, and this summer it raised two hundred million dollars at a two billion dollar valuation. Its customers are companies like CVS and Wealthfront. I couldn't find an election product.



Corn

Two billion dollars and it's being used to ask pretend people about pharmacy loyalty cards.



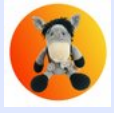
Herman

The market research money is where the revenue is. Elections are once every few years and they're graded in public.



Corn

Somebody must have been brave enough to be graded in public.



Herman

Aaru. A startup founded by two nineteen year old dropouts and a fifteen year old. In June twenty twenty-four they ran a chatbot poll of a New York congressional primary, the Bowman and Latimer race, and came within three hundred and seventy-one votes of the result.



Corn

Hm. That's genuinely impressive.



Herman

Then in November they forecast the presidential election. They had Harris winning narrowly. Trump won, and Semafor, who'd been covering them, wrote that they got most of their state calls wrong. One of the founders said they were within the margin of error, so we did well.



Corn

Every forecaster in history has wanted to say that sentence and most of them had the shame not to.



Herman

There's also Electric Twin in Britain, co-founded by Ben Warner, a former Downing Street adviser, which sells synthetic audiences. There was a study this summer showing its synthetic sample landed within the range of seven human online samples, which is interesting, except Electric Twin funded the study and both authors hold shares in the company.



Corn

The independent study was independent of everyone except the company.



Herman

If Daniel wants to build one himself, the open source option is Expected Parrot's tool, which comes out of MIT. It lets you write a survey and run it on a panel of AI agents with defined personas, across different models, and it's MIT licensed. ElectionSim, an academic project out of Fudan and Indiana, also put its code up. It draws a pool of around a million simulated voters from social media profiles and benchmarks itself against polls.



Corn

Drawn from social media. So the simulated electorate is the people who post.



Herman

Which is a real problem, and it's an old one. Ten or fifteen years ago there was a whole fashion for predicting elections from Twitter. Tweet share predicts vote share. It looked brilliant on one German election until someone pointed out that by that logic the Pirate Party should have won.



Corn

The loudest people online voting with their keyboards.



Herman

Now, the society simulations. Daniel asked whether they're relevant. Smallville had twenty-five agents in a little town who, famously, organised a Valentine's Day party on their own. Tsinghua's AgentSociety ran over ten thousand agents and about five million interactions, testing polarisation, how inflammatory messages spread, universal basic income and hurricanes. OASIS, from the CAMEL-AI group, simulates something like X or Reddit with up to a million agents and reproduces polarisation and herd behaviour. Project Sid put up to a thousand agents into Minecraft, where they voted on tax rates and spread a parody religion.



Corn

Have any of them predicted a real election?



Herman

None that I could find. Not one.



Corn

Then I'll make the case that they shouldn't. They're simulating how opinion moves. An election here is decided by turnout in particular towns, and whether one small list clears three and a quarter percent. That's not a conversation. That's arithmetic plus who bothered to get on a bus.



Herman

I half agree. Where they'd be useful is the campaign, not the count. If you wanted to know how a rumour spreads through a closed community in the last week, that's an agent simulation question.



Corn

And the wargaming connection? Because that's where I'd expect the lessons to come from.



Herman

That's where the most useful one comes from, actually. When we did that long run on AI wargaming, the thing we kept running into was evaluation. You can't really grade a war simulation, because there's only one real war and you don't get to replay it. The best known study there, from a team with a Stanford policy brief behind it, put five language models into a wargame and every one of them escalated. The raw base model, without safety training, was the most aggressive.



Corn

You never find out whether a wargame was right.



Herman

But an election is the opposite. It is the best benchmark these simulators could possibly ask for. Fixed date, public result, down to the vote. And that's why I'm a little suspicious that almost none of these companies publish an election forecast in advance. The one that did, Aaru, got graded and lost.



Corn

There's one forecaster we haven't talked about that is graded every single minute. The betting markets.



Herman

Prediction markets, and that's where a lot of the actual AI is running, quietly. Polymarket publishes its own open source framework for building AI trading agents, and there are bots that trade election markets around the clock. In twenty twenty-four the markets were more bullish on Trump than the models were, and they were right, though one French trader alone had put more than twenty-eight million dollars on him across four accounts.



Corn

One man's opinion with a very large wallet.



Herman

That's the critique. The Israel Democracy Institute cites an analysis that the top one percent of wallets held about two thirds of the volume in some markets, and that bets under a thousand dollars could move a price by ten points. On Polymarket right now Eisenkot is around fifty percent to be the next prime minister. Do you treat that as information or as noise?



Corn

Information about what people with crypto wallets think. Which is a sample, just a strange one. I'd put it next to the polls, not above them.



Herman

That's about where I left it in our forecast. A comparison, not an input. So I can hardly argue.



Corn

What about here, though? Has anyone built an AI forecast of Tuesday?



Herman

Not that I can find, no. Nobody has published a synthetic voter forecast of this election, or of the last one. What we have instead is three things. An Israeli startup called Chatoptic took twenty-six voter personas and ran them through ChatGPT, Gemini, Perplexity, Grok and Claude, and recorded which parties the chatbots recommended to them. For an ultra-Orthodox man worried about the cost of living, ChatGPT leaned Shas and Perplexity leaned Likud.



Corn

That's not a forecast. That's chatbots canvassing.



Herman

They say so themselves. The second thing is a Geocartography survey for the Israel Internet Association that found about a quarter of Israelis say they might use AI to help decide their vote.



Corn

A quarter. So the chatbots aren't predicting the election, they're in it.



Herman

Which is the more interesting story. And the third thing is the Israel Democracy Institute. On the eighth of October they proposed treating two things legally as polls under the Elections Law. Polymarket's prices, and AI synthetic simulations. That would mean the same publication ban from the Friday before the vote, plus disclosure and warning labels.



Corn

Wait. They want to ban the robots during the blackout.



Herman

They want a published simulation to be held to the same rules as a published poll, because to a voter reading a headline they look the same. And honestly, I think that's right. If a chatbot says Likud on sixty-one the weekend before an election, nobody reading it is going to check whether it interviewed anyone.



Corn

What would you actually use one of these for, if you were forecasting Tuesday?



Herman

As a stress test. I'd use a persona panel to ask questions a pollster can't afford to ask five hundred people. How does a Haredi voter in Beit Shemesh react if Shas's leader says something on Sunday. Then I'd throw the answer away as a number and keep it as a hypothesis. And the forecast itself would still start from polls.



Corn

That's a long way round to saying the polls win.



Herman

The polls with a decent model on top win. The chatbot is a scenario machine. Corn, you've got that face.



Corn

I'm just thinking that Hebrew is a small language on the internet, and the Haredi and Arab towns are the places with the thinnest online footprint. The very voters Israeli pollsters already struggle with are the ones the models know least about.



Herman

That's my guess too. The gaps would line up instead of cancelling out.



Hilbert

Synthetic. You both keep saying synthetic voters. Synthetic is a shirt you regret in August. What you're describing are index cards.



Corn

Index cards with a language model attached.



Hilbert

Esther Hazan did it with index cards. The Talpiot Model Railway Society, nineteen eighty-six. Forty-two members, and a dispute about the gauge of the main layout that had been going on for eleven years. The big gauge against the small gauge. There was an election for chairman and Esther was the treasurer, and she also took the minutes, so she knew what everybody had said at every meeting since the dispute started.



Hilbert

She wrote a card for each member. Name, gauge, grudges, whether they'd come if it rained. Every evening she laid the cards out on her kitchen table in Talpiot and ran the election. The cards voted by the society's rules, including the rule that a member who had resigned and rejoined voted twice, which nobody could remember the reason for. Turning index cards over is not easy with claws, so I pushed them with the side of my hand. I helped on Tuesdays. Corn, your chair's creaking, sit still for a second... and she'd tally them in pencil, and then update the cards from that week's minutes. She brought a tin of ma'amoul every Tuesday, date ones, even though I was small gauge and she knew it.



Herman

That's a persona panel with a memory update. That's more rigorous than half the startups I just described.



Hilbert

It was very rigorous. The problem started when the members found out about the cards. Mordechai Elbaz heard his card said he'd abstain, and he was so offended he joined the big gauge faction out of spite. Then three others asked to see their cards and started doing what the cards said, because they assumed Esther knew something. By the spring the society was behaving more like the cards than the cards were behaving like the society. The secretary moved the vote to a Thursday to stop her updating them.



Hilbert

Then the cards got a subcommittee. Mordechai said if the cards were going to decide things they should be represented, so the society minuted a motion recognising the cards as observers. Not voting members. Observers. Esther had to bring the shoebox to meetings and put it on its own chair, and Avraham Peretz, the vice chairman, would address questions to the shoebox when he disagreed with someone, so he wouldn't have to say it to their face. Twice the shoebox was asked to leave during a closed session about the clubhouse lease. Esther took it out to the corridor and sat with it.



Corn

Observer status for a shoebox is more than the Knesset gives most lists that miss the threshold.



Hilbert

It had a better attendance record than the vice chairman, I'll say that.



Corn

So the forecast started running the election.



Hilbert

The forecast won by eleven votes. The main layout went big gauge. Esther lost the chairmanship by one vote. Hers. She voted for the other man because her card said she would.



Corn

That's the argument for what the Democracy Institute wants, honestly. Once people know the simulation exists, they start voting with it or against it.



Herman

The Mordechai Elbaz problem. A published forecast changes the thing it's forecasting. A poll has the same issue, but at least a poll asked somebody.



Corn

Where I come out is that the useful question isn't whether a model can predict Tuesday. It's what happens when a quarter of the country asks a chatbot how to vote, and the chatbot has an index card on them.



Herman

And whether the rest of us find out before the Friday blackout or after. Thanks to Hilbert Flumingtop, who produced this one.



Corn

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



Herman

We'll be back soon.