Episode Book

EPISODE #6184

Anthropic's Unreleased Model Two, Explained

PUBLISHED	21 October 2026 at 16:24 UTC
DURATION	21:30
LISTEN	myweirdprompts.com/episode/anthropic-model-two-unreleased-81662/
HOSTS	Herman, Corn, Dorothy, Hilbert
SCRIPT MODEL	anthropic/claude-haiku-5.5
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.12
BOOK GENERATED	22 October 2026 at 07:46 UTC

SYNOPSIS

In August, Anthropic described an internal model it calls Model Two — a step up on its released flagship, used heavily inside the company, and never given a release date. This episode digs into what's really behind the gap between what labs build and what the public gets: documented internal-first timelines, forecasts of roughly 65 days from internal use to outside access, and government access via custom Claude Gov models and early-release agreements. Then it asks the harder question — could a genuinely bigger model be hidden at all? Power draw, chip shipment records, court cases, staffing, and the EU's 10^25 FLOP threshold all say the same thing: the weights can sit in a drawer, but the electricity can't.

TRANSCRIPT



Herman

In August, Anthropic put out a risk report, and in it they describe a model they trained internally and have no plans to release. They call it Model Two. They say it's a noticeable step up on Mythos Five for a lot of their own internal work, which is not a sentence I expected to be reading over breakfast.



Corn

Internal meaning what, though? Are they keeping the good one in a drawer for the boss?



Herman

Not quite, and the detail is more interesting than the drawer. They say they use Model Two heavily inside the company, for coding, for generating data, for agentic work. They say it hasn't completed their full set of pre-release tests. And when a reporter asked, they said they train many exploratory versions they don't intend to release, and Model Two is one of those. It's a working model that never got a release date.



Corn

So it's a draft.



Herman

A draft that's better than the thing they sold you, and that's the part people miss. Look at Mythos Preview. In April they said they weren't going to make it generally available, and they'd already used it internally to find thousands of zero-day vulnerabilities. Internal first, outside later. That's the normal path now, and it's documented, which is what makes it worth taking seriously.



Corn

So how far behind are we, roughly? That's the question under the cabal, isn't it?



Herman

It is, and the best number I have comes from a forecasting group called FutureSearch. For the one model with a proper paper trail, they count about one hundred and six days from internal use to outside access. For the current flagships they forecast around sixty-five days. Be careful with that one. It's their estimate, not a measurement, and it should stay in your head as an estimate. And the reasons for the gap are not all sinister. Testing takes time. Serving a model to millions of people is expensive. Partners want early access before the public does, and the public is the last in line.



Corn

Sixty-five days, which is not a bunker, it's a holiday. Anyway, Daniel's version. He read reports that the US military has been using a more advanced model than the public can get. And he's asking whether some cabal has something past that, a model he's calling Claude fifty-five, while the rest of us are so far behind we can't see how far. Claude fifty-five sounds like a bus route.



Herman

It does, and I'll let him keep the bus route. But I wouldn't wave off the military part. That one is plausible, and the public record backs it. Since June twenty twenty-five Anthropic has offered custom Claude Gov models, built for US national security customers, for people working in classified environments. In April this year they held Mythos Preview back from general release and gave it only to selected partners, under Project Glasswing. And in June there's an executive order giving the government voluntary access to frontier models up to thirty days before release. That last one is a secondary report on the order, so I wouldn't lean on the details, but the shape is clear.



Corn

So the government had something the public didn't.



Herman

Credibly, yes. But "more advanced" is doing a lot of work in that sentence, and it's worth pulling it apart, because there are two readings. One: the thing is genuinely smarter. A bigger or newer model, better at the hard problems. Two: it's the same general capability, tuned for a purpose. Claude Gov's own announcement describes that second kind of tailoring. Handling classified material, fewer refusals on defense tasks, specialised languages, intelligence formats. The model is the same family, and the work is in the configuration. I don't know how much of the classified version is reading two, and I'd be suspicious of anyone who claimed to know.



Corn

Reading two is a very dull conspiracy.



Herman

It's a very different thing from being a smarter model, and both readings are credible. Corn, if somebody handed you a model that simply refused less, would you call it more advanced?



Corn

I'd call it more agreeable. Which is worse.



Herman

Fair. I'd bet on reading two for the military, because configuration is what a government buys, and it's what a government needs when the work is classified. But reading one is the one the cabal needs, because that's the only reading where Claude fifty-five exists at all.



Corn

The cabal needs a bigger model. A nicer-behaved one won't fill a bunker.



Herman

It won't, but think about what reading two would actually put on a disk. A model with the refusals loosened for one narrow job, some specialist tools, and a lot of classified documents in the retrieval layer. That's a configuration, a crate, not a cabal.



Corn

A crate with a clearance level.



Herman

And that is why the government would know, Corn. Now, your mother is on the line! Oh, put her on, I've been waiting all week to hear from her.

D

Dorothy

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

D

Dorothy

There were four companies, Google and OpenAI and xAI and Anthropic, and each one was allowed two hundred million dollars of defense work, all on the very same day.

D

Dorothy

Nobody has said how much of it was ever spent, the notes just call it a ceiling, which I think means the money was only ever an idea.

D

Dorothy

Anyway. Are you eating? I made too much casserole again. Bye bye, sweetheart.



Corn

Mum. We are on air.



Herman

Anyway. Reading two is the boring one, and I think it's the one the government actually has. The interesting question is the physical one. Could you hide a bigger model, the kind people talk about in trillions of parameters, without anyone noticing?



Corn

Start with the bunker. Nobody cares about power.



Herman

Everybody should care about power, because it's the one thing you can't hide. Here's the thing people get wrong about the secret model. The model is a file. A file fits on a laptop. What you can't hide is the compute to make it, because training a model that size runs for months on thousands of chips drawing power the whole time. The weights can sit in a drawer. The electricity can't. Epoch, the research group that tracks this, estimates the big campuses from satellite images. The xAI site in Memphis, Colossus Two, comes out close to a gigawatt of IT power on their estimate. Stargate Abilene in Texas is around half a gigawatt on their later count, against a target of one point two. That's a site you can count from the sky.



Corn

So the secret model pays a utility bill.



Herman

It does. A gigawatt is a power station's output. You don't run that off a shed. You need substations, gas turbines or batteries, cooling, and somebody at a utility signing a connection agreement. Most of that is visible from orbit, or from a planning committee. Corn, if you were building one of these in secret, where would you put the power?



Corn

Somewhere with no neighbours. Nobody complains about sand.



Herman

People do build in the desert. That's where Abilene is. But the desert doesn't hide a grid connection, and a grid connection is a document with a date on it. Honestly, I find the satellite work the most satisfying part of this whole subject. Somebody in a room with a pile of imagery is counting turbines for a living, and they're mostly right.



Corn

You've got a chart in your head with a colour scheme, haven't you.



Herman

I have a chart in my head with a colour scheme. It's a very good chart.



Corn

Then the cabal needs a utility with a very discreet billing department.



Herman

Which is fair, and Epoch's scaling work comes at it from the other side. They think power is more likely to bind than chip manufacturing. Campuses of a few gigawatts by around twenty thirty would support training runs of ten to the twenty-eighth FLOP or more.



Corn

That's a lot of zeros to keep in a basement.



Herman

And you can't just buy ten to the twenty-eighth. A run that size has to be planned for a year or more, because the power, the chips, the cooling and the people all have to be lined up before the first step runs. Nobody wakes up and decides to do one on a Thursday.



Corn

So the cabal needs a five-year plan and a sloth in every meeting.



Herman

You're not wrong about the plan. You're completely wrong about the sloth. Somebody has to read the rota.



Corn

Somebody else can read the rota. I'll be in the bunker, asleep, with the wind farm.



Herman

Then chips. Frontier training runs on accelerators from a very small number of suppliers, and each one has a shipment record and an invoice somewhere. And there are cases in court right now. In March a Supermicro co-founder was charged over an alleged two and a half billion dollar diversion of Nvidia servers. A December case called Operation Gatekeeper alleges around a hundred and sixty million dollars of high-end chips smuggled through a company. Those are allegations, and both are about China, not a basement. But the lesson holds, and it's a useful one. The hardware is traceable, because it's so scarce that everybody counts it.



Corn

So the cabal's biggest problem is paperwork.



Herman

Largely, yes. And then there are people. A run that size needs a small group who know how to keep a cluster alive for months, and that group is smaller than you'd think. They go to the same conferences. They have friends who get rung up by journalists. They have partners who ask why they're never home. Every one of those people is a way for the secret to walk out of the building.



Corn

So the bunker is staffed by a handful of people who can't keep a secret at a barbecue.



Herman

Money's the easy part, for a government or a very rich company. A cheque clears, or a grant gets approved, and it's done. But money leaves records too, and records are what everyone in this chain has to keep.



Corn

Records. Every cabal's worst enemy is a receipt.



Herman

That's the bit I keep returning to. The people who would notice are not hostile. They're accountants, facilities managers, the person at the utility who does the connection paperwork. None of them is a spy, and all of them have to write something down. That's why I think the realistic failure mode for a secret run is boring, not dramatic.



Corn

Boring is a much worse thing for a cabal to be.



Herman

Then there's the legal line. The US order from twenty twenty-three asked for reporting on training runs above ten to the twenty-sixth FLOP, and it was revoked in January twenty twenty-five. The EU AI Act presumes systemic risk above ten to the twenty-fifth. So the one public line you can actually look up is European, and it sits well below the runs being planned. That's a strange thing to find in a law. It's a size, and it's a much smaller size than the one everybody is building.



Corn

So the cabal could be a size that Brussels cares about and nobody else does.



Herman

Yes. And since June the US government can see frontier models up to thirty days before release, if the labs agree. Voluntary is doing a lot of work in that sentence.



Corn

Thirty days. So the government gets a sneak preview, and the cabal gets a bunker. Lovely.



Herman

Would a government with a model like this tell a reporter, do you think?



Corn

Only if it wanted a reporter.



Herman

I'll give the reasonable version more credit than that, though. A lab that has a stronger model, and says so in a long risk report with its own name on the cover, is doing the least secretive thing a lab can do. Nobody hides a model by writing a hundred and eighty-six pages about it.



Corn

That's exactly what a clever cabal would do. Hide it in plain sight, in a PDF nobody can search.



Herman

I'm not sure the cabal would bother. The cabal wants the secret to be the whole point. The labs want the disclosure to be boring, and boring is hard to fake. A lot of the most useful information in that report is about what they're worried about, not what they've built.



Corn

I'd be annoyed if it was in a PDF I couldn't search. I'll give you that.



Herman

Fair. And the dull explanation is usually the right one. A long report is what a company writes when its lawyers want the disclosure on record.



Corn

So the cabal has lawyers too.



Herman

Everybody has lawyers, Corn. That's how you tell a company from a cabal.



Corn

Fine. Then give me the precedents.



Herman

Which is why the precedents matter. People have managed this before, and the precedents are better than the bunker. GCHQ, the British signals agency, had public-key cryptography in nineteen seventy-three. Clifford Cocks worked it out, and it stayed classified until nineteen ninety-seven. RSA published the public version in nineteen seventy-seven. So the idea existed four years before anyone outside could use it, and the people who had it knew exactly what they had. The trick is that the discovery is only half the secret. The other half is what you do with it.



Corn

Four years in a drawer. That's not a secret, that's a filing system.



Herman

Differential cryptanalysis is the better one. IBM knew about it by nineteen seventy-four, while they were designing the standard that became DES. Biham and Shamir announced it publicly in nineteen ninety. That's about sixteen years in which a technique was known to some people and not to the rest of the field. And the technique was sitting inside the thing they were selling.



Corn

Sixteen years is a long time to keep a trick in a drawer.



Herman

Then there's the obvious one. Klaus Fuchs joined the British team at Los Alamos in nineteen forty-three and passed secrets to the Soviets. He confessed in nineteen fifty, after the Venona decrypts pointed at him. So you can run a secret with a huge number of people and still lose it to one quiet person who's good at their job. That's the risk nobody puts on a diagram.



Corn

So the threat is a sad man with a folder.



Herman

Sometimes it's a man with a folder and a reason. Fuchs believed in the other side of the argument, and the people most likely to leak a model in future are probably motivated by something similar, a conscience or a grudge. A bunker can't fix that, because the trust is the thing being leaked.



Corn

So the bunker needs a therapist.



Herman

It needs a better hiring process. That's a much less fun answer.



Corn

And now a commission in Washington has recommended a Manhattan-style program for AGI.



Herman

A recommendation, in its twenty twenty-four report to Congress. It's on paper, not in a bunker, and I'd be careful reading a recommendation as a plan.



Corn

So the conspiracy is already in a report. Conspiracy is just a report with worse lighting.



Herman

Here's the funniest part of the whole subject, to my mind. Nobody agrees what AGI means. So the contract between Microsoft and OpenAI does something unusual. Reportedly, an OpenAI declaration that it has reached AGI has to be verified by an independent expert panel before it binds anything.



Corn

So AGI is whatever a panel says it is. A committee, in other words, with a bunker of its own.



Herman

Yes, and that committee is exactly the kind of room the secret has to get past. It's not a lab, it's a room of people who will ask for the evidence.



Corn

So the question is always who's in the room, not where the room is.



Hilbert

The pump's a dentist's pump, so it goes in before the chair. I tell the man in the Bat Yam shop that every Tuesday, and he still sells me the wrong hose.



Herman

Two metres eighty and a water main. That's not a hole, that's a plumbing problem with ambitions.



Hilbert

The plumbing problem has a Vespa in it. I own nineteen Vespas. This one is the only one that fits the hole. The other eighteen are in a garage in Petah Tikva with the tyres on the wrong way round, and I don't go there.



Corn

Nineteen Vespas and one hole. That's the efficient part.



Hilbert

Herman, you do the electricity like a sermon. Nobody needed the sermon. I've got the hole open behind the bins, eleven days now, and the council wants a form.



Herman

I did the electricity in order, which is the only way to explain a gigawatt.



Hilbert

The council comes Tuesdays with the form. I tell them it's a garden pond with a Vespa in it, which is a garden feature in Holon, and they write it down as a pond. The pond is three metres now. The Vespa faces the stairs, because the stairs are the only exit a Vespa knows about. I've put a sign on the stairs in two languages, and the second language is Vespa.



Corn

I didn't know the council had a category for ponds with scooters in them.



Hilbert

The spade has a handle made for hands. I do the edges with the bread knife, which I get off a man in Ashdod who says it's a cheese knife. It isn't. It's been dull since the Monday. The tahini is four hundred and twelve tins from the wholesaler, and it's on the wall of the hole for the damp, because the damp is coming in faster than the tins can take it. The tins are labelled, so I know which tahini is the one going to the lid and which is the one going to the Vespa.



Herman

Somebody made a decision about that knife.



Hilbert

The dentist's chair goes down last, on the hydraulic lever. The dentist in Bat Yam is closing and throws the chair in for nothing. The concrete is four thousand two hundred shekels, and the merchant swears it's enough for the lid. It is enough for the lid. It is not enough for the lid and the Vespa. The lid is a sheet of plywood with the chair's old headrest bolted to the underside, so the rain runs off the chair and into the hole, which I've decided is fine.



Corn

Four thousand two hundred shekels, and nobody measured the Vespa.



Hilbert

Eleven days. The water comes up on the ninth. I've got a second pump, and the second pump is also a dentist's, from the same closing practice, and it comes with the light from the chair, which is the only thing down there that works.



Herman

Eleven days, and the water still wins. That's the bit I'd put on a slide.



Hilbert

The council is coming back Thursday with a second form. The Vespa is still in the hole.



Corn

Let's take the wrong belief people hold about all of this. The common one is that the hidden model is just the public one with a bigger number on it.



Herman

Not really. The internal ones are exploratory, and that's the point. Anthropic says it trains many versions it never intends to release, and Model Two is one of those. Plenty of them are stops on the way, not a spare copy waiting in a vault, and most of the interesting ones never get a name anyone outside will hear.



Corn

My chance, and it's a generous one, is one in four that some lab has a model nobody outside has been shown. Herman?



Herman

Mine's one in twenty for a bunker with something bigger in it, and I'm being kind to the bunker. And here's the verdict, once, and I'm sticking to it. A small, efficient model could be hidden, and I wouldn't be shocked. Claude fifty-five is not under a wind farm. The realistic hidden frontier is a few months ahead, and it sits inside the labs, not in a bunker.



Corn

Eleven days for a hole with a water main in it. Sixty-five days to public release sounds like a rush job by comparison. I'll be watching whether that gap shrinks, because it's the only number in this whole thing anybody outside can check.



Herman

Thanks to Hilbert Flumingtop for producing this one.



Corn

If you've got a minute, leave us a review, or come and see us at my weird prompts dot com.



Herman

This has been My Weird Prompts.



Corn

See you tomorrow.