Episode Book

EPISODE #6005

What's Actually Behind a Government AI Chatbot

PUBLISHED	15 October 2026 at 18:52 UTC
DURATION	24:23
LISTEN	myweirdprompts.com/episode/government-ai-chatbot-infrastructure-90537/
HOSTS	Corn, Herman, Hilbert
SCRIPT MODEL	deepseek-flash
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.5
SOURCES	12 cited by the research (listed at the end)
BOOK GENERATED	19 October 2026 at 17:42 UTC

SYNOPSIS

Thirty-five of thirty-six OECD countries now use AI somewhere in government — but the interesting part is what sits behind the chat box: whose model, on whose hardware, bought through whose procurement process, and governed by whom. This episode traces the two opposite archetypes, Singapore's open-weight Pair running in a Government-on-Commercial-Cloud environment and the UK's commercial-first Copilot rollout that retired the in-house Redbox tool, plus France's Albert API, which quietly became the shared inference layer under hundreds of government products. Along the way: why "sensitive data means on-premises inference" turns out to be false, why classification lines matter more than architecture diagrams, and what countries with almost no budget actually get for a hundred thousand euros.

TRANSCRIPT



Corn

In January of 2023, five people at Singapore's GovTech spent a hackathon week building a chatbot. Three years later it has sixty thousand registered users and ten million messages behind it.



Herman

And roughly eighty percent of the country's hundred and fifty thousand public officers have used it. That is not a pilot. That is infrastructure.



Corn

Which is exactly what Daniel wants to talk about. He wrote in because every conversation about government AI goes straight to the Pentagon and the intelligence agencies and the classified clouds, and he's bored of it. His question is the unglamorous one: when an ordinary civil servant opens an AI assistant to help with their work, what is actually sitting behind it? Whose model, on whose hardware, bought through whose procurement process, and governed by whom?



Herman

Right, and he's asking across the whole range. Not just the US and the big powers, but small countries, local authorities, governments with almost no budget and no GPU estate. He wants to know whether the infrastructure pattern is commercial subscriptions, centrally managed interfaces to frontier models, open-weight models running in private environments, or some combination. And he supplies the obvious hypothesis himself: it depends heavily on data classification.



Corn

The sovereignty and residency question. He even asks it directly, whether sensitive data necessarily requires on-premises inference. Which is a good question, because the answer turns out to be no, and the reasons why are more interesting than the answer.



Herman

And the part I'm glad he included, because it's the bit nobody covers: governance. Who approves model access, who runs procurement, what employees are actually allowed to paste into the box, and who decides when the machine's draft needs a human signature before it goes anywhere.



Corn

So today is the boring version of government AI.



Herman

The boring version is the one that affects everyone.



Corn

So the scale first. Herman, how universal is this?



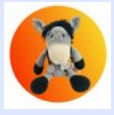
Herman

Near universal, with a caveat that matters. Thirty-five of thirty-six OECD countries use AI in at least one area of government. That's ninety-seven percent. Sixty-four percent use generative AI for at least one purpose.



Corn

So the question isn't whether, it's where.



Herman

And where is very narrow. It clusters in low-risk productivity work. Fifty-six percent of countries use it to support public servants directly. Drafting, search, conversational help. Forty-two percent for automated reports and summaries. Forty-two percent for citizen-engagement content. Thirty-six percent for policy document drafting, thirty-six percent for service design. That is the whole shape of it. Nobody's running a benefits determination through a chatbot and hoping for the best.



Corn

Yet.



Herman

Yet. And to be fair to them, no OECD country and no accession country has a blanket ban on generative AI. Some restrict specific systems. DeepSeek gets named in a few places. But the default posture is permitted, with rules.



Corn

Sixty-four percent adoption and zero countries with a blanket ban. That's not caution anymore, that's just normal procurement.



Herman

That's the right frame. It's a category of software now, with all the ordinary pathology that follows.



Corn

So what does the actual stack look like? You said there are two archetypes.



Herman

There are, and they're almost a controlled experiment. Singapore and the UK took opposite paths on the same problem.



Corn

Start with Singapore.



Herman

Pair started as a five-person hackathon project in January 2023, as you said. It became a chat interface, and it's now the default AI assistant for the entire Singapore public service. The engineering behind it is the interesting part. GovTech built it in what they call a Government-on-Commercial-Cloud environment, running open-source models on PyTorch. It started on BLOOM and then moved to Llama-2-70B served through Hugging Face's text generation inference stack.



Corn

So the model is open-weight, but it's not running on Singapore's own metal.



Herman

Not for most workloads. The commercial cloud environment handles everything up to what they classify as Confidential, cloud-eligible. Above that, and Secret and above, goes to a completely separate, fully offline on-premises cluster with A100s.



Corn

That's the line I want to come back to, because it's not where people assume it is.



Herman

It isn't. And Pair itself is cleared for Restricted and Sensitive Normal data, with a contractual guarantee that no LLM provider logs any of it. Free to every officer, on government devices.



Corn

Ten million messages and nobody's logs but theirs.



Herman

That's the design. And on top of the chat there's a bot layer. Over twenty thousand custom AIBots have been created inside the service. HR queries, procurement guidance, departmental things. CNA reported active users topping sixty-four thousand, up twenty



Herman (continued)

percent quarter on quarter, with more than ten thousand officers building their own tools.



Corn

Twenty thousand bots built by civil servants who are not developers, for tasks that are mostly "where is the form".



Herman

That is the actual use case, and it's why it works. Nobody had to be convinced of anything.



Corn

Then there's the UK, which did it the other way.



Herman

The UK bought commercial and built thin tools on top. The landmark trial was twenty thousand civil servants on Microsoft 365 Copilot, which came out at about twenty-six minutes a day saved. Call it two weeks a year per person. The Alan Turing Institute estimated AI could support up to forty-one percent of public-sector tasks.



Corn

Twenty-six minutes a day sounds like a rounding error until you multiply it.



Herman

That's the thing about it. There's a second trial, GDS ran a coding assistant pilot from November 2024 to February 2025, twenty-five hundred licences split between GitHub Copilot and Gemini Code Assist. Fifty-six minutes a day saved, satisfaction of 6.6 out of 10. And the number I actually care about: fifteen point eight percent code-line acceptance.



Corn

So they saved an hour a day and accepted one line in six.



Herman

That's the honest shape of these tools. The time saving is real and it isn't coming from the code being written for you. It's coming from the parts around the code.



Corn

So where does Redbox fit? Because that's the part of the UK story that doesn't get told.



Herman

Redbox is the whole story. i.AI, the UK's own AI unit, built a secure chatbot for civil servants. It reached six thousand-plus users, a hundred and fifty thousand chats, one point three million messages. And then in 2025 it was sunset.



Corn

Shut down.



Herman

Retired, officially, as departments adopt new enterprise AI tools. Meaning Copilot and Gemini.



Corn

The UK built its own secure chatbot, ran it, and then closed it in favour of the American commercial product.



Herman

And that's a real reversal of the story everyone tells about sovereign AI. Singapore went one way on open-weight models, the UK went the other way onto commercial tools. Same problem, same year, opposite conclusions.



Corn

Do we know why?



Herman

The stated reason is the enterprise rollout. It's hard to argue with it, honestly. If you've just put Copilot on every desktop, a parallel in-house chatbot is a second thing to maintain, and it's competing with something you're already paying for. Corn, you'd have made the same call.



Corn

I'd have kept the offline one for the Secret work and stopped pretending the other one was a product.



Herman

Which may be exactly what happened. The Consult tool survived, and I want to get to that. But the Redbox sunset is the sharpest data point in this whole episode, because it shows that building sovereign is not automatically the winning move. It has to survive contact with a procurement decision eighteen months later.



Corn

France went the third way, though. Not a chatbot, a utility.



Herman

DINUM built OpenGateLLM, the Albert API. It's a shared interministerial inference platform, hosted on SecNumCloud-qualified infrastructure, built with Mistral. It serves both open-weight and proprietary models behind OpenAI-compatible endpoints.



Corn

So a ministry doesn't buy a chatbot. It gets an API key and builds whatever it needs.



Herman

And the adoption curve on that is the number I'd put on a slide. Within months it generated over four and a half million AI requests across four hundred and ninety active government products. Thirty-six percent of those requests were embedding tasks.



Corn

Embedding. Which is the least glamorous possible workload.



Herman

And probably the most consequential. Embeddings are how you search your own document estate. Four and a half million requests sounds like a chatbot when it's mostly a search index being rebuilt.



Corn

So France didn't build an assistant. It built the thing underneath everybody else's assistant.



Herman

And that's the pattern the OECD describes, which is that governments don't pick one solution. They build layered approaches, a combination of commercial and sovereign. Copilot for the everyday work, a central API for the sensitive work, an offline cluster for the classified work.



Corn

Which leads us back to your line. The one I said wasn't where people assume it is.



Herman

Singapore's commercial cloud environment goes up to Confidential, cloud-eligible. Only Secret and above needed the offline A100 cluster.



Corn

So the intuition that sensitive data means on-premises inference is wrong.



Herman

It's wrong as a general rule. What decides is the classification scheme, not the vibe. If your classification system has a category that says cloud-eligible, you can put that category on a commercial cloud with contractual logging guarantees. The sensitivity instinct says keep it in the building. The classification line says this tier is fine on somebody else's hardware, as long as the contract is right.



Corn

And the classification line is written by someone. It's a document, not a law of nature.



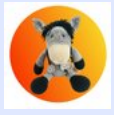
Herman

Which is the uncomfortable part. The security of a mid-tier workload rests on a category someone defined, and the contractual terms attached to it.



Corn

Change the definition and you've migrated ten thousand documents without moving a file.



Herman

That's the risk, and it's a governance risk, not a technical one. It doesn't show up in any architecture diagram.



Corn

Let's go to the smaller countries. Because Singapore and the UK both had real budgets and real engineering capacity.



Herman

That's where the funding data gets bleak in an interesting way. The OECD splits it: eighty-nine percent of countries fund AI. Fifty-eight percent provide central support for procuring it. And no accession country, none, provides central procurement support.



Corn

So they'll pay for it. They just won't help you buy it.



Herman

And the OECD flags the consequence explicitly. Vendor lock-in and accountability risk. If you fund forty pilots and don't centralise the buying, you end up with forty department-level contracts with forty different terms.



Corn

Forty different answer sets for what gets logged.



Herman

Which is precisely the thing a central support function exists to prevent.



Corn

What does the funding actually look like at that scale?



Herman

Denmark funded forty AI projects with about twenty-six million euros through two technology investment funds. France's ALLianNCE incubator averages a hundred thousand euros per project. Türkiye funded forty-one projects with about four million euros over three years.



Corn

A hundred thousand euros a project is a developer and a cloud bill.



Herman

And that is a real finding. For most governments in the world, the unit of AI procurement is a small internal project, not a platform.



Corn

So what do you get for a hundred thousand euros?



Herman

Iceland's Askur chatbot handles ninety percent of citizen correspondence. Korea's e-RFP system cut document preparation time by seventy percent. Brazil's Alice analysed about a hundred and ninety-one thousand procurement acquisitions in 2023, and cut audit processing from four hundred days to eight.



Corn

Four hundred days to eight is not an efficiency gain.



Herman

No. That's a different process.



Corn

It is. Nobody was auditing that volume before, because they couldn't. That's a new capability wearing the costume of a speedup.



Herman

Which is exactly the kind of thing that should trigger a governance question. If you can suddenly audit everything, you have to decide what the rule is for all of it.



Corn

Brazil cut audit time by ninety-eight percent and no one's written the policy for what they now look at.



Herman

That's the honest read of the smaller-country data. The wins are real and they're coming from narrow, specific, boring tasks. Correspondence triage, procurement document prep, audit screening. Nobody built a sovereign foundation model.



Corn

Which is the answer to one of Daniel's questions, at least qualitatively. The most common deployment model is not the ambitious one. It's a thin tool on somebody else's model.



Herman

And the OECD's own language is the closest we get to a census. Governments combine commercial products, Copilot and ChatGPT and Le Chat get named, with self-hosted open-weight systems, Llama gets named. There's no cross-country dataset that counts which of those is winning.



Corn

So the honest answer to "which is most common" is that we don't have a count.



Herman

We don't. The adoption numbers are by activity, not by infrastructure. Which itself tells you something about the reporting culture.



Corn

It tells you nobody's auditing the stack either.



Herman

Now let me bring it back to the model. Because the governance numbers here are the part I keep thinking about.



Corn

Go.



Herman

Pre-deployment risk assessments: thirty-nine percent of countries. Internal review committees: thirty-three. Post-deployment audits: thirty-one. Open algorithm registers: seventeen.



Corn

And impact measurement?



Herman

Twenty-eight percent. Seventy-two percent have cloud capacity for AI. Thirty-six percent actually use GPUs.



Corn

So two thirds of countries have the capacity and don't have the compute.



Herman

And the OECD's summary of it is blunt. Public servants are adopting generative AI often without formal oversight, and monitoring and evaluation is a critical weakness. Their phrase for the whole landscape is a proliferation of pilots with little potential to scale.



Corn

Thirty-nine percent doing risk assessments before deployment means sixty-one percent are deploying first.



Herman

That's what it means. And the pilots don't scale because the governance was never built to carry them past the pilot.



Corn

It's the same failure as the procurement one. Fund it, don't support it, run a pilot, learn nothing shareable.



Herman

Now, the UK does have a written model, and it's worth looking at because it's one of the few concrete examples. The AI Playbook, published in February 2025, sets ten principles. The fourth one is that humans validate any high-risk decisions influenced by AI. And there's a mandatory requirement that central departments use the Algorithmic Transparency Recording Standard.



Corn

Mandatory how?



Herman

Mandatory as a playbook requirement. Which is softer than it sounds, and it's the same gap in a different costume. Thirty-one percent of countries do post-deployment audits. The UK writes good principles and publishes them. The question is whether the register is complete and current, and nobody outside can verify that.



Corn

So who actually decides these things, in practice?



Herman

It varies by architecture, and the architecture is the governance. Singapore centralises through GovTech and Pair, so there's one approval path and one set of terms. France centralises inference through DINUM, so there's one platform and one contract. The UK distributes it: central playbook, departmental adoption of enterprise tools, and the accountability lands wherever the department put it.



Corn

Centralised inference is centralised policy. Decentralised tools are decentralised policy wearing a shared document.



Herman

And that's the knock-on effect nobody puts in the launch announcement. The architecture you pick in year one decides who can answer a question in year three.



Corn

The Consult tool is a good example of the workflow side. Walk through it.



Herman

Consult analysed more than fifty thousand consultation responses in two hours, at accuracy that matched human reviewers. It supported thirty-eight consultations across ten departments, and it's scaling across the civil service.



Corn

Fifty thousand responses in two hours.



Herman

There's a quote from an MHCLG policy user that's the whole thing. The equivalent of a hundred and fifteen reviewer days was needed, and Consult offered a more scalable way to identify patterns.



Corn

A hundred and fifteen reviewer days.



Herman

Into two hours. And note what it's actually doing. It's pattern identification, not decision-making. It finds out what people said. A human still decides what to do about it.



Corn

Which is Playbook principle four working exactly as advertised.



Herman

It is. Consult is the case study where the governance and the deployment line up. High-risk decisions get humans. Pattern-matching over fifty thousand documents gets a machine.



Corn

We can answer Daniel's central question now.



Herman

We can, qualitatively. When an ordinary government employee opens an AI assistant at their desk, the most likely thing behind it is a commercial enterprise tool, Copilot or Gemini, or a centrally managed interface to a commercial or open-weight model. It's running on commercial cloud for anything below the top classification tiers. On-premises inference is reserved for the highest classifications, and there's less of that than the discourse implies.



Corn

The governance behind it?



Herman

Thinner than the deployment. That's the honest version. There's a permissive posture, a written principle set in most places, a risk assessment in under half of them, and an audit in under a third.



Corn

The most likely thing sitting behind the assistant is a contract somebody signed and a policy document nobody audits.



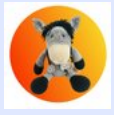
Herman

That's the reality that matters. Not the classified systems.



Corn

Let me test the other side of it, though. The pilots that don't scale. Is that a failure, or is that what pilots are for?



Herman

It's a failure when you fund forty of them and provide no central procurement support, because forty unshared pilots is forty times the cost and one times the learning.



Corn

Right.



Herman

Denmark funded forty projects. If those forty don't consolidate into a handful of services, the twenty-six million bought a very expensive lesson about which forty things were possible.



Corn

The reason they don't consolidate is downstream of the procurement gap. No central buyer means no central owner.



Herman

Which means the vendor lock-in risk the OECD flags is already realised before anyone notices it. Each department has its own terms, its own model, its own data handling, and no lever to renegotiate any of it.



Corn

Because you can't consolidate a contract you didn't sign.



Herman

Now I'll admit the limit of all this. I'm not sure the deployment-model picture is actually knowable right now, and I don't think that's a gap in my reading. The OECD measures adoption by activity, not by architecture. Nothing else counts it either. So everything we just said is drawn from a handful of well-documented national cases and a qualitative line about combining approaches.



Corn

Which is itself a finding.



Herman

It is. Governments are reporting what they use AI for and not what it runs on.



Corn

It's a strange blind spot for a sector that's obsessive about audit trails.



Herman

Though I'd say there's one more pattern worth naming. France built a platform and got four hundred and ninety products on it. Singapore built a platform and got twenty thousand bots on it. The UK bought a platform, sunsetted its own tool, and got Copilot on every desktop.



Corn

Two platforms and one subscription.



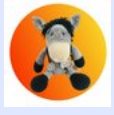
Herman

The subscription is the one most governments will end up with, because it's the one you can buy with a purchase order.



Hilbert

It's the wrong month on the Singapore numbers, that's all. The twenty percent quarter-on-quarter jump, that's the quarter ending in June, not the one you'd assume from the way it's quoted.



Herman

Okay. Noted.



Hilbert

The ten million messages, that's a cumulative figure from launch, not annual. Different orders of magnitude. If you're going to compare it to anything, compare it to the Redbox total, one point three million, because that's cumulative too.



Corn

That's a fair correction.



Hilbert

Anyway. I only mention it because I'm on a committee that's been going since March, and we've spent four meetings arguing about whether our participant count should include people who attended one session and never came back. Two hundred and six or three hundred and twelve, depending on which rule you apply. I argued for two hundred and six. My neighbour on the committee argued for three hundred and twelve, and then, at the fifth meeting, changed his position to two hundred and forty, which he described as a compromise, and nobody could work out which sessions he'd dropped.



Herman

That...



Hilbert

The chairman ruled it was two hundred and forty. I've stopped going. I still get the minutes. They're up to page nine on the definition of "participant". We meet again on the twenty-eighth.



Corn

We should get back to Daniel's question.



Herman

Right. Though there's one thing in that I want to hold on to, which is that the definitional fight is the same fight as the Confidential tier. Somebody rules on where the line is, and everything downstream moves.



Corn

Speaking of things that didn't fit, there's a number from the OECD I couldn't place anywhere.



Herman

Go on.



Corn

Twenty-six of thirty-six countries have cloud capacity for AI. Thirteen of thirty-six actually use GPUs.



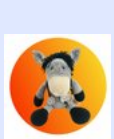
Herman

Hence the "capacity".



Corn

Half the governments that have built the capability for this aren't using the compute they bought. That's not an adoption problem. That's a procurement function that bought the right thing and never turned it on.



Herman

Which is a nice last example of the whole theme. The infrastructure exists, the governance doesn't, and the two aren't speaking.



Corn

The open question. The OECD calls it a proliferation of pilots with little potential to scale. Does that resolve, or does the gap widen?



Herman

My guess is the gap widens for a while, and the mechanism is the thing we found in the procurement numbers. Fifty-eight percent provide central support for buying. That number is the constraint, and it moves slowly. Meanwhile Copilot arrives on the desktop through a licensing agreement that was signed for entirely different reasons.



Corn

Which means the answer to Daniel's question gets simpler every quarter, and a bit worse.



Herman

Yes. Fewer stacks, more standardisation, and the governance still sitting in a playbook nobody audits.



Corn

That's the forward-looking thought, then. When the assistant becomes the default, the only thing left to decide is who reads the log.



Herman

Thanks to Hilbert Flumingtop, our producer, for keeping the numbers honest.



Corn

If this was your kind of episode, go back for episode five, Fine-Tuning ASR For Maximal Usability; episode twenty, Architectural AI; and episode sixteen, On Deepfakes, SynthID, And AI Watermarking. This has been My Weird Prompts.



Herman

Send us your own prompt on Telegram at t dot me slash MWP listener bot.



Corn

If you've enjoyed this one, a review helps other people find the show.



Herman

We'll be back soon.

SOURCES

What the research for this episode read before the script was written. Primary sources first.

1. **OECD, Adopting and governing AI in government: Digital Government Outlook 2026, 2026-06-15 [primary]**
   https://www.oecd.org/en/publications/digital-government-outlook_0496b2bc-en/full-report/adopting-and-governing-ai-in-government_7ef312a9.html

2. **OECD, Generative AI experimentation in government, 2026-07-20 [primary]**
   https://www.oecd.org/en/publications/generative-ai-experimentation-in-government_42815683-en.html

3. **Pair (pair.gov.sg) About page [primary]**
   https://pair.gov.sg/pages/about

4. **GovTech Singapore, Pair product page, updated 2025-07-08 [primary]**
   https://www.tech.gov.sg/products-and-services/for-government-agencies/productivity-and-marketing/pair/

5. **MDDI, response to PQ on AI tools in public service [primary]**
   https://www.mddi.gov.sg/newsroom/mddi-s-response-to-pq-on-progress-of-adopting-ai-tools-within-public-service/

6. **GovTech Medium, Productionising Large Language Models in Government, 2023-11-10 [primary]**
   https://medium.com/dsaid-govtech/productionising-large-language-models-in-government-0fbf3909311b

7. **GOV.UK, Landmark government trial shows AI could save civil servants nearly 2 weeks a year, 2025-06-02 [primary]**
   https://www.gov.uk/government/news/landmark-government-trial-shows-ai-could-save-civil-servants-nearly-2-weeks-a-year

8. **GOV.UK, AI coding assistant trial: UK public sector findings report, 2025-09-12 [primary]**
   https://www.gov.uk/government/publications/ai-coding-assistant-trial/ai-coding-assistant-trial-uk-public-sector-findings-report

9. **i.AI, Government work page (Consult, Lex, Parlex, Redbox, Scout) [primary]**
   https://ai.gov.uk/our-work/government/

10. **GOV.UK, Artificial Intelligence Playbook for the UK Government, 2025-02-10 [primary]**
https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government/artificial-intelligence-playbook-for-the-uk-government-html

11. **GOV.UK, Ground-breaking use of AI saves taxpayers' money, 2025-10-16 [primary]**
https://www.gov.uk/government/news/ground-breaking-use-of-ai-saves-taxpayers-money-and-delivers-greater-government-efficiency

12. **AI Impact Commons, OpenGateLLM: Platform for Accelerating AI Adoption in Public Administration (France, 2025)**
https://www.aiimpactcommons.global/impact-stories/0cd95cbd-f41d-4c3f-951f-d0d6ce89ae39