



Who actually holds the controls

© John Stroh

Version 0.2 · revised 8 September 2026 · [this version as a PDF](#)

An institution that cannot say who governs its AI agent has not bought a tool. Five questions, the answers the market currently gives to them, and the Māori data governance work that made the underlying argument first.

The pursuit of ‘Goodness’ in AI — Sovereign agents, and what it takes to trust one with real work

What a contract settles, and what it leaves

FIG-27 FIVE GATES, AND WHO CAN OPEN EACH

THE QUESTION	WHO CAN ANSWER IT
1 Whose agent is it? Which named body holds the authority, can be asked for its reasons, and carries the consequences.	no vendor can answer this one
2 What authority does it hold? Purposes, data, actions, prohibitions, expiry — in a form a board member can recite.	a limit stated only in the agent's instructions is a limit that argument can move
3 Where does its operational life take place? Not where the records rest — where inference, retrieval, logs, credentials and support reach from.	
4 Who may inspect and challenge it? Both halves are required. The second is routinely absent.	
5 Who may decommission it, and with whose agreement?	Te Kāhui Raraunga's Māori AI Governance Framework names this a partnership decision

Ordered so that a failure on any one of them is visible without technical knowledge. The fifth question is not this series' own.

5 An institution that cannot answer the first has not delegated to the agent. It has installed it.

FIG-27 The five questions as a sequence rather than a checklist, or any one is visible without technical knowledge. The right-hand column is the part that decides the answer: the first question is one no vendor can answer, and an institution that cannot answer it has installed an agent rather than delegated to one.

 **Ask the Glossary**

Ask an institution running an AI agent whose agent it is, and the first answer will be a contract. The institution pays for the service, the terms name the institution as the customer, and in every ordinary commercial sense the agent is the institution's. That answer is true and it settles almost nothing, because the question is not who bought the thing. It is who governs it: who set the limits it operates under, who may inspect what it did, who may change what it is permitted to do, and who may stop it.

Those are different questions from ownership and they have different answers, which is easiest to see by noticing that all four can be answered by somebody other than the purchaser without any term of the contract being breached. The supplier changes the underlying model and the agent's behaviour changes. The supplier's logging tells the institution what the agent did in the form the supplier chose to record it. The supplier's terms permit the supplier to withdraw a capability on notice. Nothing there is a breach. The institution has bought a service and acquired no authority over it.

This piece sets out five questions that separate the two, examines the answers currently being sold, and identifies the body that put the same argument in writing before this series did.

What the market means by the word

The commercial definitions are worth reading closely, because they are careful, they are not dishonest, and they answer a different question from the one an institution needs answered.

NVIDIA's is the most widely circulated. Sovereign AI, on its own public page, is **“a nation's domestic capability to produce artificial intelligence using its own infrastructure, data, workforce, and business networks”**, the goal being **“to protect and preserve local languages, values, culture, and history”**. Read it for what it contains: infrastructure, data, workforce, networks — a definition of *productive capacity*. Then read it for what it does not contain. There is no term in it for who may inspect the system, no term for who may change what it is permitted to do, and no term for who may stop it. A nation could satisfy that definition completely and hold no authority over the thing it had produced.

Microsoft's is more granular and, in one respect, more useful, because it concedes something. Microsoft Sovereign Cloud is described as **“a suite of capabilities and deployment models designed to help governments and regulated industries meet stringent data residency, compliance, and operational sovereignty requirements without sacrificing the benefits of hyperscale cloud innovation”**, offered in three tiers. Of those tiers, Microsoft's own documentation says: **“Azure local and private clouds provide the strongest sovereignty controls by offering full control over hardware, software, data, location, and management. However, those environments don't deliver the full cloud**

value.” The sentence is an accurate description of a real trade-off, and it establishes that even the vendor treats sovereignty as something bought at a price rather than supplied as standard. ⚠️ The page carries a July 2026 date and is post-cutoff.

AWS defines the territory by its commitments rather than by a sentence: control over where data sits, **“additional access restrictions that limit all access to customer data unless requested by the customer or a partner they trust”**, encryption, and the ability to **“sustain operations through disruption or disconnection”**. Again the axis is control over assets, and again there is no term for authority.

None of those three definitions is a lie, and it is worth being precise about that, because the argument here does not depend on anyone having lied. Each answers the question its author is in a position to answer. A supplier can sell location, encryption, access restriction and resilience. It cannot sell an institution the standing to govern, because standing is not a property the supplier holds and could transfer.

Two pieces of evidence from the supplier's own record

The strongest evidence that residency is not authority is published by the companies selling residency, and no adversarial reading is required to find it.

Microsoft's Azure Local **disconnected operations** is the most sovereign offering in that catalogue: the control plane runs locally, on a dedicated management cluster, and the documentation states that **“when you run disconnected, data, operations, and control remain within your organization's boundaries.”** That is precisely the arrangement an institution seeking independence would want. The eligibility terms for it, on the same documentation, are these. **“You need an eligible agreement with Microsoft.” “You must have an active support plan... with Microsoft.” “You need a valid business need for running and operating in a disconnected environment.”** Access requires a prequalification form and Microsoft's approval, stated as arriving within ten business days. The arrangement is licensed on a capacity-based model with an annual term, billed monthly.

Set those two halves against each other. The control plane is local. The authority to *have* a local control plane is granted by the supplier, conditioned on a commercial agreement, subject to the supplier's assessment of whether the institution's need is valid, and renewed annually. An institution running that arrangement has genuine technical control and holds it at another party's discretion. ⚠️ Post-cutoff; the pages carry 2026 dates and rest on retrieval.

The second piece of evidence is sworn. On 10 June 2025 the French Senate's commission of inquiry into public procurement asked Microsoft France's director of public and legal affairs, Anton Carniaux, whether he could guarantee under oath that French citizens' data held by Microsoft would never be transmitted following an injunction from the United States gov-

ernment without the explicit agreement of the French authorities. His answer, in the Senate's own record: **“Non, je ne peux pas le garantir, mais, encore une fois, cela ne s’est encore jamais produit.”** — No, I cannot guarantee it, but, once again, that has never yet happened.

Both halves of that sentence belong wherever it is quoted. The qualifier is on the record and it matters: no such transfer is claimed to have occurred. What the answer establishes is not misconduct but structure. Data held in France, under a French public procurement contract, with every residency assurance in place, remains reachable by a legal process the French institution is not party to and cannot resist. The residency was real. It was not authority, and the person best placed to know was asked under oath to guarantee otherwise and could not.

Where this argument was already made

The argument that authority rather than location is the test was not reached by this series independently, and presenting it as an original finding would be both wrong and weaker than the truth.

Te Kāhui Raraunga published the **Māori Data Governance Model** — *Tuia te korowai o Hine-Raraunga* — in 2023, authored by Kukutai, Campbell-Kamariera, Mead, Mikaere, Moses, Whitehead and Cormack, and structured around eight pou. Its statement of the point is direct: Māori data sovereignty **“extends beyond mainstream concepts of data sovereignty which are primarily concerned with data residency and jurisdiction.”** The banner under which the whole model sits is **“Māori authority over Māori data”** — the object named there is authority.

The 2025 **Māori Artificial Intelligence Governance Framework**, contributed to by Chris Cormack, Erena Mikaere and Te Taka Keegan, carries that into AI and states the control-plane argument in terms this series can only cite rather than improve on:

There are also risks involved when data, being used by AI, is stored onshore using a global cloud service provider. In most instances, its staff in various jurisdictions abroad will be able to access the data, network and storage configuration details, and have hypervisor access. Both the USA and China assert jurisdiction over data stored by companies headquartered in their respective countries.

And on the same page: **“Māori authority over data regardless of storage location or AI processing jurisdiction.”** The framework goes further than this series does in one respect worth recording rather than softening — it states that **“the use of Generative AI tools built on Māori data but not built by Māori should be prohibited!”** — and it names decommissioning as a decision requiring partnership, which is the fifth of the five questions below.

Two things follow. The first is that an argument this series would otherwise be advancing on its own authority has been made, in an Aotearoa context, by a body acting for iwi, hapū and whānau Māori, and citing them is both accurate and stronger than not. The second is a caution about a related error. Te Mana Raraunga — the Māori Data Sovereignty Network — remains active, and held a members' hui at the University of Waikato in 2025 to set priorities for the coming decade. Dr Karaitiana Taiuru's critical analysis of 20 September 2025 concerns that network's **2016–18 principles**, which he argues do not reach AI bias, training sets, predictive policing or digital colonialism, and which he records as “**largely not implemented in practice**” despite academic influence. A critique of a document is not a critique of a body, and this series does not conflate the two.

The agent nobody chose

Everything above assumes an institution chose an agent, which is not how most of them meet one. The ordinary route is that an AI feature appears inside a product already being paid for — the accounting package, the job-management system, the customer database, the email suite. Nobody evaluated it. Nobody signed for it. It arrived in a release note, and by the time anyone considered whether to use it, staff had been using it for a month.

Put the five questions to an arrangement of that kind and the answers are not merely poor. They are structurally unavailable.

Whose agent is it? Nobody in the organisation granted it anything. The authority under which it operates is whatever the terms of the original product allow, and those terms were agreed for a different product, before this capability existed.

What authority does it hold? Whatever the supplier's feature does, which changes when the supplier changes it. There is no grant to inspect, because no grant was made.

Where does its operational life take place? In most cases the customer cannot find out. The feature may call a model the supplier licenses from somebody else, in a jurisdiction disclosed nowhere in the contract.

Who may inspect and challenge it? The supplier's logs show what the supplier chose to record. A customer disputing what the feature did about them is disputing it with an organisation that has no visibility of the mechanism.

Who may stop it, or move it? Frequently there is a toggle, and frequently the toggle governs the interface rather than the processing. Leaving means leaving the underlying product, which is the thing the organisation actually depends on.

An institution in that position has not made a bad decision about AI. **It has not made a decision about AI**, and the acts of the system are attributed to it regardless.

The practical consequence is a question worth putting to any supplier, in writing:

which of your features use a model, whose model is it, where does it run, what does it retain, and what would we have to do to turn it off without losing the product?

The five questions

An institution can establish whether it governs its agent, rather than merely paying for it, by requiring answers to five questions. They are ordered so that a failure on any one of them is visible without technical knowledge.

Whose agent is it? Which named body — not which individual, and not which job title — holds the authority under which this agent acts, can be asked afterwards for its reasons, and would carry the consequences? A vendor cannot supply this answer, and an institution that cannot supply it has not delegated to the agent; it has merely installed it.

What authority does it hold? Which purposes, which data, which actions, which prohibitions, which expiry — stated in a form a board member can recite. The test is not whether a document exists. It is whether the limits it states are enforced somewhere other than in the agent's instructions, because a limit expressed to a system as a request is a limit that argument can move.

Where does its operational life take place? Not merely where the records rest, which is the question residency answers. Where the inference happens, where the retrieval happens, where the logs are held, where the credentials are issued and revoked, and where support access reaches from. Te Kāhui Raraunga's framing is the useful one here: the question is what the governing institution's authority actually reaches, and the answer is frequently narrower than the map suggests.

Who may inspect and challenge it? Both halves are required, and the second is routinely absent. A system may be perfectly explicable to a person who has no route to change the record the explanation rests on, and that arrangement is a wall with good signage.

Who may stop it, or move it? Whether the institution can revoke the agent's credentials, export its records in a form readable without the supplier's software, move its functions elsewhere, and retire it — without requiring the supplier's cooperation and without the supplier retaining practical control of anything that matters. This is the question the market answers least well.

An agent for which all five have answers is what this series calls a sovereign agent. The term is not about hardware location and not about scale. A small charity running an agent on a leased server it does not own may satisfy all five; a government running one in its own building may satisfy none.

What the instruments require, and what they leave out

It would be an overstatement to say the questions are unaddressed. Three instruments address parts of them, and reading what each covers shows precisely where the gap sits.

The **EU AI Act** requires, at Article 14(4), that a person overseeing a high-risk system be able to “**decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output**” and to “**intervene in the operation... or interrupt the system through a ‘stop’ button or a similar procedure that allows the system to come to a halt in a safe state.**” Article 26(2) requires deployers to “**assign human oversight to natural persons who have the necessary competence, training and authority.**” That is a genuine legal answer to part of the fifth question — but it governs the relationship between the overseer and the system, not between the institution and the vendor. A stop button that halts a system the supplier can restart, or withdraw, or alter, answers a smaller question than the one asked here.

NIST’s AI Risk Management Framework asks organisations to establish mechanisms to supersede, disengage or deactivate systems performing inconsistently with intended use. Again the assumption is that the deploying organisation possesses the technical ability; the framework has no occasion to ask whether a third party could countermand it.

The **EU Data Act**, at Article 23, comes closest to the fifth question and is binding. It requires providers to remove “**pre-commercial, commercial, technical, contractual and organisational obstacles**” to a customer terminating a contract, contracting with a different provider, porting exportable data and digital assets **including to on-premises infrastructure**, and achieving functional equivalence. That is a legal right to leave, and it is the single most useful existing lever an institution has. It concerns data portability rather than agent governance, and the AI Act does not cross-reference it — but a European institution asking who may move its agent has a statutory answer that an American one does not.

What no instrument asks is the first question. Nothing retrieved requires an AI agent to hold an identity of its own, distinguishable from the person on whose behalf it acts. The nearest thing is a control that exists and was then left optional: NIST SP 800-53 **IA-9** requires that system services and applications be uniquely identified and authenticated before communicating — and it appears in no baseline. Not Low, not Moderate, not High, not Privacy, not the operational-technology overlay. An organisation bound by those baselines is not required to select it. The control was written, and then made optional, and that is a more interesting fact about the state of the field than an absence would have been.

Two questions each organisation answers for itself

How much authority is it reasonable to hold at another party's discretion? Almost every institution depends on infrastructure it does not control, and a standard that admitted no dependency would exclude every organisation on earth, which means it is the wrong standard. The useful question is not whether dependencies exist but whether the institution knows which ones it could not survive the withdrawal of, and has said so out loud.

What is the institution prepared to give up to answer the five questions? Microsoft's documentation states the trade-off plainly, and it is not a trick: the strongest sovereignty controls do not deliver the full value of the hyperscale platform. An institution that wants both is not being cautious; it is refusing to decide. The choice is real and it belongs to the governing body rather than to whoever is configuring the system.

How you would know this is wrong

First, if a supplier can be shown to transfer governing authority rather than capability — if a contract exists under which the institution, not the vendor, decides what the model may be changed to, when a capability may be withdrawn, and what the record must contain — then the distinction drawn here between buying and governing is weaker than this piece claims, and the five questions collapse into ordinary procurement.

Second, if institutions holding none of the five answers turn out to govern their agents effectively in practice — because commercial pressure, professional norms or regulation supply what the arrangement does not — then the argument is over-specified and the questions are an audit nobody needs.

Third, if the identity requirement turns out to exist somewhere this search did not reach, the claim about IA-9 being the high-water mark is wrong and should be corrected publicly rather than quietly. Two searches were run for it, both keyword-based over standards and regulatory corpora, which means they shared an assumption and do not constitute two independent methods. **The finding is therefore: no requirement was found. It is not: none exists.** Anyone who knows of one is invited to say so through the contact route on [the about page](#), and this piece will be corrected.

Related

[What has to be settled first](#) sets out why these five are asked at all.

The six, in reading order. Each stands on its own; read together they build one argument.

- What has to be settled first — why the prior questions are prior, and what each of the others answers
- How much it may do unsupervised — five levels of permitted autonomy, declared rather than left implicit
- The four properties underneath — four properties a delegation must carry, each checkable by observation
- What you can actually require — twelve requirements, and what currently requires each of them
- What nobody has measured yet — four open questions, and the observations that would settle them
- An invitation to become a Distributor — not part of the argument above: a proposal to work with its author, for anyone who wants to take this further

ACKNOWLEDGEMENT

Leslie Stroh — my big brother, a product philosopher, and a magnificent role model. He helped me find my way to a vision of goodness in AI.

►DISCLAIMER

How this was made. The version number counts drafts of the text. It does not measure the inquiry behind it, which has run over days and across several AI systems, with argument between those systems and within them, directed, refused and repeatedly redirected by the author. The source material was AI-generated, and then adversarially and iteratively refined across a range of tools — systems built by different companies in different jurisdictions, set against each other and against the author. No one of them produced this text, and no one of them reviewed it alone. **The plurality is deliberate rather than incidental.** A single model carries a single set of priors about which sources are authoritative, and this series argues that an evidence base narrowed in exactly that way is how a contested question comes to look settled. Using one model to investigate that claim would have been the claim refuting itself. To name a single model on it would credit that model with work that was neither its own nor done in a single pass. The diagram on each page was drawn afterwards, with the same assistance, from the argument the essay had already made. **The plurality was also necessary, and the record should say why.** In drafting, the assisting model repeatedly led with United States institutional sources — a national laboratory, an industry association, a market study nineteen years old — and presented conclusions drawn from them as the state of knowledge. On one occasion European measured data contradicting those conclusions was present in the same research return and was placed below them. Framings were proposed that would have argued against this series' own position using that evidence base, and offered as rigour. Each was refused by the author and the material rebuilt. That is the mechanism these documents describe, occurring in their own making, and it is recorded because a series arguing that evidence bases narrow without anyone deciding to narrow them cannot credibly claim its own production was exempt. The framing, the corrections and the judgements are the author's, and so are the errors.

Cite this version

Who actually holds the controls, version 0.2 (September 2026).
agenticgovernance.digital/downloads/whose-agent-is-it-v0-2.pdf

Licensed CC BY 4.0. Reuse is free, including commercially, provided the author is credited and the reuse does not suggest that the author endorses it. This link is frozen at version 0.2; [/whose-agent-is-it](#) always shows the current one.

PREVIOUS

What has to be settled first

Series contents

NEXT

How much it may do unsupervised

Alongside: [questions and answers](#) · [sources and provenance](#) · [slides](#)

Share this

Mastodon

Email

Copy link

Found something wrong?

This piece names what would show it is wrong. If you have that — a figure, a source, an argument — it is worth more to us than agreement. Tell us which claim and why.

Which section, if you know it

for example: 2. Four properties

The claim you are responding to

quote it, or describe it

What is wrong with it

Your email, only if you want an answer

Used to reply to you and nothing else. No list, no newsletter, never passed on.

Send this