Federate, Don't Align

The safe path through the AI sovereignty contest — why federated communities, and federated inference, are the lowest-risk option for nations that will never win the capacity race

Précis. A small nation is usually told its AI future is a choice of allegiance — the American stack or the Chinese stack. Read instead as a risk decision, three options appear, and they fail differently. Aligning with either superpower's stack buys capability at the price of a dependency you cannot reverse on your own timetable: one is built around a registry you do not control, the other around a centre with no exit. Federating — holding your own data and models, and reaching others' through signed, consent-bound, instantly-revocable envelopes that carry inference as well as records — accepts a capacity ceiling and coordination overhead, but carries no irreversible tail. It is sovereignty as rightful authority rather than raw capacity: the layer a community can hold today without winning a race it was never going to win. A federated Aotearoa — iwi, hauora, education, and research communities composing national-scale capability with no national registry and no foreign substrate — shows the shape of it. The safe option is not the powerful one; it is the only one whose worst case you can walk back.

A small nation deciding how to meet the age of agentic AI is usually told it has two doors. Behind one is the American stack — frontier models, hyperscaler clouds, the commercial frontier. Behind the other is the Chinese stack — state-coordinated models, a national agent fabric, an industrial policy that ships. The debate is conducted as a choice of allegiance: which superpower's AI do you build your public institutions on? I want to argue that this is the wrong question, and that asking it as a question of *risk* rather than allegiance changes the answer. There is a third door, it is already open, and for most of the world it is the safe one.

A companion piece, *The Map Has No Node for Legitimacy*, makes the underlying case in full: that "sovereignty" runs together two different things — the *capacity* to build and run AI without depending on anyone, and the *rightful authority* to decide how a system behaves on the people it acts upon — and that the strategic-competition frame measures only the first. I will not re-argue that

here. I will take it as given and ask the practical question it leaves open: if a small nation accepts that it will never win the capacity race, what is the lowest-risk way for it to hold the authority it actually can?

The choice as a risk decision

Strip away the allegiance framing and three options remain, each with a different risk profile. The point is not that one is virtuous and the others are not. It is that they fail differently, and one of them fails in a way you can recover from.

Align with the American stack. You get capability now, and you accept a set of standing exposures: your public data flows through infrastructure governed by another country's law, including its extraterritorial reach; your dependency deepens with every integration, so the cost of leaving rises over time; and your continuity is hostage to commercial decisions and policy weather you do not control — a model deprecated, a price changed, an export rule rewritten. None of these is catastrophic on any given day. The risk is that they are not reversible on your timetable.

Align with the Chinese stack. The People's Republic of China's 2026 *Implementation Guidelines for Intelligent Agents* describe a coherent and, on its own terms, impressive architecture: a national *intelligent internet* with a central agent-registration platform that gives every agent a queryable identity; categorised and tiered governance, in which the state and its industry authorities decide which classes of application get direct scrutiny; and rule-embedding with behavioural fencing, constraints baked into agents and verified centrally. It is sovereignty as capacity, executed well. But it is built around a single substrate that authority flows down through, and it has no exit. For a community that is not the state operating that fabric, the standing exposure is total: the registry that makes your agents legible is the registry that makes them governable by someone else.

Federate, and align with neither. Hold your own data and your own models, and reach other communities' data and models through bilateral, consent-bound, instantly-revocable channels. You accept two real costs: a capacity ceiling — your models are smaller, your substrate is borrowed — and coordination overhead, because federation is many agreements rather than one platform. What you do not accept is an irreversible tail. There is no central point whose compromise reaches you, no dependency you cannot withdraw from, no authority above you that cannot be refused. Every link is one you can cut, with your records and their provenance intact on the way out.

The safe option is not the powerful one. It is the only one of the three whose worst case you can walk back.

The mechanism: federation of communities, and of their inference

The reason this is more than a slogan is that the unit of federation can be made small, signed, and revocable — and that the same channel can carry both data and AI.

In the architecture this paper draws on, two installations exchange information only through a *signed envelope*: consent-bound, scope-limited, addressed to a named recipient, and carrying the cryptographic provenance of every record inside it. There is no shared platform that a third party pivots through, and no central authority that can revoke a participant's standing. Withdrawal is immediate and exit is without penalty.

The move that matters for AI is that the same envelope carries *inference*. A community's situated language model — a model trained on that community's own content, under its own authority — can reach a peer community's corpus through the envelope, scoped to a fixed query shape, and answer against it without that corpus ever leaving the peer's control or its provenance being stripped. This is federated inference: capability assembled by composition across consenting holders, rather than concentrated in one model that everyone must query and therefore trust.

Set this beside the central-registry design and the difference is structural, not ideological. A national agent platform is efficient precisely because it is one thing — which is also why its compromise, capture, or policy reversal reaches everyone attached to it at once. A federation has no such centre to seize. The cost of that resilience is the coordination overhead named above. The benefit is that there is no single key whose loss is everyone's loss.

This is what "rightful authority" looks like once it is made mechanical rather than asserted. Authority over how a model behaves on your data is not a permission the platform grants you and could revoke; it is a property of the envelope and the signed record, held by you, enforced by refusal. When a community withdraws its steering, the model does not fall back to a vendor's default — it falls back to silence on that domain. Refusal, not substitution, is the guarantee.

The authority layer — and a name for it

It is worth being exact about what federation does and does not buy, because the honesty is the credibility. It does not close the substrate gap. The model weights, the accelerators, the compute beneath all of this sit inside exactly the foreign-controlled levers the capacity model maps; a federated community runs open models it did not author, on hardware it did not build. That dependency is real and is not waved away.

What federation buys is the separability of two things that are usually assumed

to travel together: governance sovereignty and substrate sovereignty. The first — authority over data, provenance, steering, and the behaviour of a model acting in your name — can be held today, in full, by actors who will never hold the second. Capability at this layer is gained by *composition*: a federation of modest models, each sovereign over its own domain, reaching each other under consent, can do at national scale what no one of them could do alone — without any of them surrendering authority to a centre.

There is an older word for declining to resolve a bipolar contest by joining one side of it. In the twentieth century, nations that refused to be sorted into the American or Soviet bloc called themselves *non-aligned* — not neutral, not passive, but organised around the refusal itself, and stronger together for it. A non-aligned layer for AI is the same posture rendered in architecture: small nations and indigenous polities that decline the US-or-China sorting, gain capability by federating with each other rather than aligning above themselves, and hold authority at the layer where authority can actually be held. The point of the name is not rhetoric. It is to mark that this is a recognised category of strategic behaviour with a track record, not a novelty — and that it is available now, to actors the capacity race has already written off.

A federated Aotearoa

Make it concrete in one small country. Aotearoa New Zealand has roughly five million people and no hyperscaler. On the capacity axis it does not register, and it never will. On the authority axis it has something most large states lack: a constitutional settlement that already assumes plural, co-equal centres of authority rather than one.

Picture the mesh rather than the platform. An iwi runs a Village holding its own records and a model trained to its kaupapa. A hauora service runs another, governing clinical and whānau-ora data under its own authority. A kura or a wānanga runs a third; a research group, a fourth; te-reo infrastructure like Papa Reo, language and tooling projects like Kete AI, Māori Lab, kahu.code each run their own. None of them holds a frontier model. Each federates with the others through bilateral, scope-limited, revocable envelopes — sharing data where consent is given, and consulting each other's situated models where it helps. The aggregate is national-scale capability with *no national registry* to capture and *no foreign substrate* in the production path: inference on sovereign New Zealand infrastructure, the governance framework and federation protocol released into the commons under a reciprocal licence so any local team can fork and extend them.

The governance form this needs is not a single sovereign but a polycentric one — many co-equal authorities with distinct jurisdictions over a shared resource, in Elinor Ostrom's sense. That is not an imported abstraction here; it is the shape of the institutional landscape already. And the reason a mesh can hold together without collapsing its members' values into one number is the subject

of a separate argument — that living organisations decide in *kōrero*, holding plural commitments in tension rather than scalarising them, a point where Isaiah Berlin's value-pluralism and Christopher Alexander's pattern language turn out to describe the same thing a hui has always done. Federation is the technical expression of that: many authorities, composed, none collapsed.

The trust that makes it safe is carried in the record itself. On every sovereign record, kaitiakitanga is not a label or a policy setting but a cryptographically signed field — who holds guardianship, under what tikanga the record was shared, bound to a provenance hash that survives federation and cannot be quietly altered by an operator. Authority does not have to be trusted to a centre because it is carried, signed, in the thing itself. This is not a roadmap: demonstration communities of several types are running today, with seeded data, as working proof that the envelope, the situated model, and the signed record compose end to end.

What this is, and what it is not

Federation does not win the capacity race; it refuses to enter it. It leaves substrate dependence exactly where it is and does not pretend otherwise. It is more work than buying a platform, and it asks communities to hold authority rather than delegate it. These are real costs, and a serious account names them.

What it offers in return is the one property the two alignment doors cannot: no catastrophic, unrecoverable bet. A federated community can lose a partner, a model, even a substrate provider, and recover — because it never handed anyone the single key. For a small nation, and for any community deciding whether to pour its public life into infrastructure it cannot govern, that is not a consolation prize. It is the definition of the safe option. The capacity contest was never one most of the world could win. The authority layer was always open. The federated mesh is where a great deal of the world will end up living — and it is already running.

The threads this synthesises

This argument stands on a body of published work. Each piece carries one load-bearing idea; stated atomically, they are the threads this paper draws together.

The Map Has No Node for Legitimacy is the diagnosis. It argues that the dominant model of AI sovereignty — capability measured across accelerators, datasets, workforce, electricity, and water — quietly equates sovereignty with capacity, and so has no term for legitimacy or for the authority of the governed. It separates effective control (de facto) from rightful authority (de jure), and shows a community can hold the second without the first. The present paper is its operational sequel: where to hold that authority, and at what risk.

Sovereign-Record Architecture (Paper A, v4) is the mechanism. It sets out a design in which the sovereignty a community needs is a property of the records

themselves — cryptographic provenance, kaitiaki attribution, and a federation envelope signed onto each record — rather than a concession an operator may revoke. Its §7 documents bilateral federation already running in production: the signed, consent-bound, scope-limited channel this paper treats as the way out.

Held in kōrero, not collapsed to a number is why a federation can hold together. It argues that living organisations decide by holding plural commitments in tension rather than reducing them to a single optimisation score, and that Isaiah Berlin's value-pluralism and Christopher Alexander's pattern language describe, in Western terms, what a hui has always done. That is the reason a polycentric mesh need not collapse its members' values into one.

A Civil-Society Proposal for Sovereign and Federated Agentic AI in Aotearoa New Zealand (v1.2) is the policy form. It mirrors the People's Republic of China's 2026 *Implementation Guidelines for Intelligent Agents* point for point, proposing the same governance surface built on sovereign records and bilateral federation rather than central registration — a constructive parallel, not an opposition.

AI Policy in China summarises those Guidelines: a national intelligent-internet with central agent registration, categorised and tiered governance, and rule-embedding with behavioural fencing. It is sovereignty as capacity, coherently executed — and the foil this paper measures the federated option against.

Source materials

Published research — agenticgovernance.digital

- The Map Has No Node for Legitimacy — the diagnosis this paper builds on.
- Sovereign-Record Architecture for Community-Scale Platforms (Paper A, v4) — §7, bilateral federation in production.
- Held in kōrero, not collapsed to a number — plural values, living organisations, and AI
- A Civil-Society Proposal for Sovereign and Federated Agentic AI in Aotearoa New Zealand (v1.2)
- AI Policy in China — a summary of the 2026 Implementation Guidelines for Intelligent Agents
- Sovereignty Isn't Only Something You Can Buy — the legitimacy argument in short form.

Demonstration Villages — mysovereignty.digital *(seeded data, in active development; brief scheduled maintenance on the hour)*

- All demonstration Villages
- Kāhui Māori · Whānau · Community · Membership

External works

- Clancy, T. & Naugle, A. B. (2026). *AI Sovereignty: A Qualitative Model of Strategic Competition as AI Becomes an Instrument of National Power.* arXiv:2606.07245.
- Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press.
- Te Kāhui Raraunga (2023). *Māori Data Governance Model* (with the CARE Principles for Indigenous Data Governance, 2020).
- Cyberspace Administration of China (2026). *Implementation Guidelines for Intelligent Agents.*

The Village platform and the Tractatus framework are an attempt to make AI sovereignty achievable for actors who will never win the capacity race, by relocating sovereignty to the layer where rightful authority can actually be held — and by letting those actors federate that authority rather than surrender it.